CDF/MEMO/STATISTICS/PUBLIC/5928
Version 2.10
June 24, 2005

Objective Bayesian Upper Limits for Poisson Processes

Luc Demortier
Laboratory of Experimental High-Energy Physics
The Rockefeller University

"...all animals, including non-Bayesian statisticians, are at least informal Bayesians."
I.J. Good [1]

Abstract

We discuss the Bayesian approach to the problem of extracting upper limits on the rate of a Poisson process in the presence of uncertainties on acceptance and background parameters. In single-channel searches, we show that the usual choice of prior leads to a divergent posterior density from which no upper limit can be extracted. We propose a solution to this problem, that involves making more efficient use of the information contained in the data. We then generalize this solution to multiple-channel searches, and describe an importance sampling Monte Carlo method to perform the corresponding multi-dimensional integration. The frequentist properties of the proposed method are also studied.

1 Introduction

CDF techniques to set upper limits on the rate of a Poisson process often use a hybrid of frequentist, Bayesian, and likelihood ideas, combined with various approximations to take systematic uncertainties into account. A prime example of such a technique is provided by the 1991 CDF paper on a search for top quarks in the electron+jets channel [2]. There are several nuisance parameters in this analysis: background, efficiency, energy scale, luminosity, etc., but they are not all eliminated in the same way. For instance, the likelihood is *maximized* with respect to the background, but *integrated* with respect to the efficiency. Furthermore, after all the nuisance parameters have been eliminated, upper limits on the $t\bar{t}$ production cross section $\sigma_{t\bar{t}}$ are obtained by treating the likelihood function as a probability distribution for $\sigma_{t\bar{t}}$.

There are two types of problem with such an approach. The first one is how to interpret the final result: what does the quoted upper limit really mean? Does it have good frequency properties? Is it plausible? The second issue concerns the mathematical properties of this procedure. Does it always converge and give sensible and robust results? If not, is it "physically" clear why?

The Bayesian framework provides a unified, consistent way to solve the problem of upper limits in the presence of nuisance parameters. Although not devoid of ambiguity[1], it is usually straightforward to apply and interpret. We provide a brief summary of this paradigm in section 2. The problem of setting upper limits in single-channel searches is described in section 3, with special emphasis on the choice of prior. The "usual" choice of prior, flat in signal cross section and Gaussian in acceptance, leads to an improper posterior density from which no upper limit can be extracted. We show how this difficulty can be avoided by a careful study of the information flow from the data to the unknown parameters of the problem. The solution is then generalized to multiple-channel searches in section 4. Section 5 is devoted to the numerical implementation of this solution via Monte Carlo importance sampling. Finally, section 6 studies the frequentist properties of this Bayesian method. Technical details are relegated to an appendix.

2 Summary of the Bayesian paradigm

Suppose we have observed some data x with probability density function (pdf) $f(x \mid \theta, \nu)$, where θ is a parameter of interest and ν is a nuisance parameter, i.e. a parameter that is not interesting but is required in order to draw some inference about θ from x. Let $\mathcal{L}(\theta, \nu \mid x)$ be the likelihood function of the measurement. By definition, the likelihood is proportional to the pdf $f(x \mid \theta, \nu)$, considered as a function of (θ, ν) with x held fixed at its observed value. The proportionality factor can depend on x but *not* on the parameters (θ, ν); it is irrelevant for inference about the latter and is therefore not further specified.

In Bayesian statistics, θ and ν are viewed as random variables, not in the sense that their true values can fluctuate, but rather that information about them can be represented coherently by a probability distribution. It is therefore possible to introduce a *prior* density for the parameters, $\pi(\theta, \nu)$, which summarizes our knowledge about (θ, ν) *prior* to the measurement x. Bayes' theorem then specifies how the prior information

[1] The main element of controversy in the Bayesian paradigm is the choice of prior (see section 2.1), especially when there is little or no prior information available about one or more parameters. The corresponding element of controversy in frequentist statistics involves the specification of the reference ensemble of measurements in which the actually performed measurement is considered to be "embedded" for purposes of interpretation. Faced with controversy in both methodologies, some authors argue that the Bayesian paradigm is to be preferred because it is *coherent*; see [3] for a discussion.

about (θ, ν) is updated by the measurement of x to yield the *posterior* density of (θ, ν):

$$p(\theta, \nu \,|\, x) \;=\; \mathcal{L}(\theta, \nu \,|\, x)\, \pi(\theta, \nu) \,/\, p(x), \tag{2.1}$$

where $p(x)$ is the marginal probability density of the data, sometimes also called predictive density:

$$p(x) \;=\; \int d\nu \int d\theta\, \mathcal{L}(\theta, \nu \,|\, x)\, \pi(\theta, \nu). \tag{2.2}$$

It is independent of the parameters and is therefore only needed as a normalization factor when one is calculating actual posterior probabilities. To obtain a posterior probability density for θ only, the rules of probability theory specify that the nuisance parameter ν must be integrated out from the joint posterior for θ and ν:

$$p(\theta \,|\, x) \;=\; \int d\nu\, p(\theta, \nu \,|\, x). \tag{2.3}$$

The Bayesian solution to the problem of extracting all available information about θ from the data is given by this marginal posterior density $p(\theta \,|\, x)$. It is then up to the user to summarize this information by providing the posterior mean or mode, credibility intervals, upper limits, etc. The ability to summarize posterior information this way is a great advantage enjoyed by the Bayesian statistician. In contrast, the frequentist statistician must decide *ahead* of the measurement whether he intends to report an upper limit, a two-sided interval, or a significance, if he wishes to avoid biasing his result.[4] Of course, in order to be useful, the posterior $p(\theta \,|\, x)$ must be *proper*, i.e. normalizable, which requires that $p(x)$ from equation (2.2) be finite.

We end this section with a comment about notation: throughout this note the greek letter π indicates prior probability densities and the roman letter p posterior or predictive densities. We do not attempt to distinguish explicitly between the various functional forms a prior or posterior density can take. The argument list should provide all the context needed to make this distinction.

2.1 Subjective vs. Objective priors

The above summary glossed over the important but difficult issue of choosing a prior density for the parameters of a problem. Depending on how much information is available prior to the measurement, there are two general approaches, labeled "subjective" and "objective". In a typical example from high-energy physics, that of the measurement of the cross section of a new physics process, one would follow the subjective approach to construct priors for nuisance parameters such as the acceptance, the integrated luminosity, and the expected background, whereas an objective method would be chosen to represent prior ignorance about the cross section parameter.

In the subjective approach, it is assumed that prior information about a parameter can be obtained from auxiliary measurement results, Monte Carlo studies, theoretical considerations, or reasonable beliefs. Often such information is summarized by a central value together with an uncertainty. These two numbers are then *interpreted* as the mean and width of the corresponding prior density, whose shape is taken to be Gaussian, log-normal, gamma, or another plausible function. Clearly there is subjectivity involved in this construction, and Bayesian robustness studies must be performed in order to determine the sensitivity of posterior inferences to prior assumptions. In general one can expect the influence of the prior to diminish as the size of the data sample increases, but the actual level of influence may need to be quantified in specific applications.

On the other hand, objective methods are used when there is no prior information, or one wishes to pretend that this is the case. This leads to the somewhat controversial concept of ignorance priors, also called variously "noninformative", "reference", "non-subjective", and "objective" priors. The form of these priors is determined by a formal rule. Some possible rules are:

1. Insufficient reason
   If the parameter space is discrete and finite, and there is lack of sufficient reason for assigning higher probability to some parameter values than to others, then one is led to a uniform prior. A natural generalization is to assign a uniform prior to continuous parameter spaces.

2. Invariance
   Changing the units of a measurement can be considered as a joint transformation of parameter space and sample space. The likelihood is invariant under such a transformation, and it may seem reasonable to require the same invariance in the prior. More general transformation groups can also be used in this argument.

3. Coverage matching
   Given some interval for a parameter, one may require the posterior probability content of that interval to match its frequentist coverage. This can usually not be achieved exactly, but only to some limited order of approximation in the sample size n. Furthermore, such matching requirements do not necessarily fix the form of the prior completely. For example, in a multi-parameter problem, requiring matching to $\mathcal{O}(n^{-1})$ fixes the form of the prior with respect to the parameter of interest but not with respect to the nuisance parameters. Coverage matching priors are noninformative to the extent that by allowing a frequentist interpretation of posterior intervals, they "let the data speak for themselves."

4. Maximal missing information
   The gain in information provided by an experiment can be quantified with the so-called Kullback-Leibler distance between the prior and posterior densities. It is possible to choose a prior that maximizes this distance in some sense and therefore acts as a "reference" prior, against which the effect of informative (subjective) priors can be evaluated.

Further details on these rules, as well as additional rules, can be found in reference [5]. Very often objective priors are also improper, i.e. their normalization is infinite. As a result, they cannot be interpreted in terms of standard probabilities (satisfying the Kolmogorov axioms), and the derivation of Bayes' theorem must be generalized. This can be done, but at the price of creating foundational difficulties such as marginalization paradoxes, incoherence, etc. There is also no longer any guarantee that the posterior will be proper. While there is no easy solution to this problem, the derivation of objective priors from formal rules provides a compelling rationale in many applications and often leads to sensible results. However, when it does not, one should keep in mind that the cause of the problem and its eventual resolution must rest with the improper prior.

E.T. Jaynes has argued that some of the foundational problems with improper priors can be avoided by only considering those improper priors which can be approached as well-defined limits of sequences of proper priors [6, Chapter 15]. Consider for example a Gaussian likelihood with unit variance and unknown mean θ, and the improper uniform prior $\pi(\theta) = 1$. The corresponding posterior is proper and can also be obtained by using the proper prior:

$$\begin{aligned} \pi_k(\theta) &= \frac{1}{2k} && \text{if } -k \leq \theta < k, \\ &= 0 && \text{otherwise,} \end{aligned}$$

and taking the limit $k \to \infty$ *after* calculating the posterior. Not all improper priors can be equivalenced in this way, but in this note we will only consider those that can. A critique of Jaynes' arguments can be found in [7].

3 Upper limits in single-channel searches

Assume we are searching for a new process, and in the absence of any compelling evidence of its manifestation, we wish to set a limit on its production cross section σ. In the simplest case we observe a single event count n and the likelihood has the form:

$$\mathcal{L}(\sigma, A \mid n) = \frac{(\sigma A L + b L)^n}{n!} e^{-\sigma A L - b L}, \tag{3.1}$$

where A is the total signal acceptance, which combines all relevant efficiencies and geometrical and kinematical acceptances, b is the expected background rate per unit integrated luminosity, and L is the total integrated luminosity for the relevant data taking period.

In practice we often do not know A, b and L exactly, but we do have information about these parameters from auxiliary measurements, Monte Carlo calculations, and/or theoretical considerations. We will show in section 3.6 that uncertainty about b does not cause any noteworthy difficulties, so for the moment we will assume that b is known exactly. For the purpose of calculating posterior probabilities, uncertainties about L are of the same nature as those about A; for simplicity we will therefore also take L to be exactly known, and we will assign a non-zero uncertainty only to A.

We now have a problem with one parameter of interest, σ, and one nuisance parameter, A. If we have some estimate of the mean value and standard deviation of A, then its prior can be represented by a Gaussian density truncated to values between 0 and 1 of its argument. On the other hand, nothing is known a priori about σ, so that some kind of noninformative prior is called for. Reference [8] lists three such priors for a Poisson distribution with mean μ: the uniform prior $\pi(\mu) = 1$, the Jeffreys prior $\pi(\mu) = 1/\sqrt{\mu}$, and the log-uniform prior $\pi(\mu) = 1/\mu$. The likelihood of equation (3.1) has the additional complication that $\mu = \sigma AL + bL$ with *both* σ and A unknown. A reasonable option is to set the σ prior to $\pi(\sigma) = \sigma^{-\gamma}$, with the values $\gamma = 0$, $1/2$, and 1 being of particular interest. The joint prior for A and σ can then be written as:

$$\pi(\sigma, A) = \pi(\sigma)\,\pi(A) \;=\; \sigma^{-\gamma}\,\vartheta(\sigma)\;\frac{e^{-\frac{1}{2}\left(\frac{A-A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\,K\,\Delta A}\;\vartheta(A)\,\vartheta(1-A), \tag{3.2}$$

with: $\quad K \equiv \frac{1}{2}\left[\mathrm{erf}\Big(\frac{A_0}{\sqrt{2}\,\Delta A}\Big) + \mathrm{erf}\Big(\frac{1-A_0}{\sqrt{2}\,\Delta A}\Big)\right] \quad$ and $\quad \mathrm{erf}(x) \equiv \frac{2}{\sqrt{\pi}}\int_0^x dt\; e^{-t^2},$

and where the step function $\vartheta(r)$ is defined to be 0 for $r < 0$, and 1 otherwise. From this prior and the likelihood, we can calculate the marginal data density, i.e. the posterior normalization factor:

$$\begin{aligned} p(n) &= \int_0^1 dA \int_0^{+\infty} d\sigma\; \mathcal{L}(\sigma, A\,|\,n)\;\pi(\sigma, A) \\ &= e^{-bL}\sum_{i=0}^{n}\frac{(bL)^i}{i!}\;\frac{\Gamma(n+1-i-\gamma)}{(n-i)!}\;\int_0^1 dA\;(AL)^{\gamma-1}\;\frac{e^{-\frac{1}{2}\left(\frac{A-A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\,K\,\Delta A}. \end{aligned} \tag{3.3}$$

For $\gamma \le 0$, the $A^{\gamma-1}$ factor in the integrand causes the integral over A to diverge at its lower limit. If γ is a strictly positive integer, the integral converges but the Gamma function in the term(s) where $i \ge n + 1 - \gamma$ diverges. In these cases the posterior density is improper and cannot be used to extract intervals or upper limits. It is interesting to note that $p(n)$ is finite for $\gamma = 1/2$; this corresponds approximately to Jeffreys' prior, which is indeed known to almost always yield proper posteriors.[8] Unfortunately, the $1/\sqrt{\sigma}$ prior has some features that may be considered undesirable. It leads to a posterior density that is always infinite at $\sigma = 0$ (although it is still integrable), and the Bayesian credibility of posterior intervals tends to overestimate their actual frequentist coverage (see section 6 and reference [9]). It is partly for this reason that a noninformative uniform prior[2] is usually preferred in High Energy Physics. This is the prior, corresponding to $\gamma = 0$, that we will consider in most of this note. Clearly the first issue we need to address is the impropriety of the resulting posterior. It is perhaps worth mentioning that physicists have often tried to use this posterior and failed to notice its divergence because of the inevitable truncation that occurs in numerical computations. As stated in [5, §4.2.5]:

> *Improper posteriors will sometimes reveal themselves by creating obvious numerical problems, but this is not always the case. Because of increased computing power, analysts use models of ever greater complexity, which in turn makes it more difficult to check whether the posterior is proper.*

[2] A flat prior, by giving equal weight to all allowed values of σ, would seem to provide a good mathematical representation of ignorance. On the other hand, it can also be argued that for any value $\sigma_0 > 0$, no matter how large, this prior always assigns more probability above σ_0 than below it. In this sense a uniform prior indicates a prior belief that the true value of σ must be quite large. This is a conservative assumption for upper limit calculations, but not necessarily for significance calculations for example.

In fact, as we will show in section 3.2, improper posteriors can fail to reveal themselves numerically even in very simple models. It is therefore always important to gain some analytical insight into a posterior density before doing numerical calculations with it.

3.1 Other aspects of the problem

When the divergence of equation (3.3) was first encountered, it was quickly discovered that changing the acceptance prior $\pi(A)$ from a truncated Gaussian to any density that is zero at $A = 0$ removes the divergence in the posterior normalization. For example, taking a beta density, $\pi(A) \propto A^{\kappa}(1-A)^{\lambda}$ with $\kappa > 0$ and $\lambda > -1$, yields a perfectly proper posterior for σ. A truncated log-normal density would also work. This type of solution has other problems however:

- It is ad hoc. Indeed, it totally ignores the fact that, from a Bayesian point of view, the real source of the problem is not the acceptance prior but the cross section prior, which is flat from 0 to $+\infty$ and hence improper. Posteriors are only guaranteed to be proper if the priors themselves are all proper. To modify the acceptance prior in order to solve a problem caused by the cross section prior smacks of "adhockery", which has no place in the Bayesian paradigm (see for example [10, §1.38]).

- It lacks robustness. Usually, all we know about the acceptance A are estimates of its mean and standard deviation. We then plug these estimates into some "reasonable" functional shape $\pi(A)$. Given *only* the very limited prior information we have about A, there is nothing that makes the beta density more reasonable than the truncated Gaussian. One simply can't claim much knowledge about the exact shape of the acceptance prior. It is therefore essential that the end-result of one's analysis not be too sensitive to that shape. One should for example be able to replace a default choice of beta by a truncated Gaussian without running into serious trouble. This issue of robustness is critical in Bayesian inference. [10, §7]

- It leads to incoherence. Suppose we do assign a beta or a truncated log-normal prior $\pi(A)$ to A. With a flat improper prior on the cross section σ, a straightforward calculation shows that the marginal posterior density for the cross section is indeed proper. On the other hand, the marginal posterior density for the signal acceptance is $p(A \mid n) \propto \pi(A)/A$, which is different from the prior $\pi(A)$ and at the same time independent of the integrated luminosity L [11]. As a result, even if we perform an experiment with zero integrated luminosity, information about the signal acceptance will be non-trivially updated. However, an experiment with zero integrated luminosity is equivalent to no experiment at all, in which case information about the signal acceptance remains unchanged. This situation, where two equivalent approaches to a problem lead to contradictory results, is a form of incoherence and is obviously undesirable.

As a counterargument to the objections of adhockery and lack of robustness, it is sometimes claimed that "the probability of zero acceptance is always finite for a truncated

Gaussian," and since one "can't set a limit if there is the possibility of no sensitivity," the truncated Gaussian is not a good model [12]. This claim is incorrect however: the probability of zero acceptance is actually zero for a truncated Gaussian. It's the *probability density* that is non-zero, and there is no obvious physics argument why it should vanish.

3.2 A plausible but ultimately unsatisfactory solution

A possibly reasonable solution would be to make the cross section prior proper by cutting it off at some (large) value $\sigma_{\max}$. This guarantees a proper posterior and removes the incoherence mentioned in the previous section, but all inferences about σ, including upper limits, will depend on the choice of $\sigma_{\max}$. This is illustrated in Figure 1 for a simple example of five events observed over a background rate of 0.02 pb, an integrated luminosity of 100 pb^{-1}, and a signal acceptance of 0.020 ± 0.006. The 95% C.L. upper limit on the cross section is plotted as a function of the prior cutoff $\sigma_{\max}$. For small $\sigma_{\max}$ values the upper limit rises linearly, with a slope equal to $\alpha = 0.95$, due to the dominance of prior information over data information in that region; it then flattens out over several orders of magnitude before reaching a steadily rising asymptotic regime. To gain some quantitative insight into this regime, we first note that an α-credibility level upper limit $\sigma_{ul,\alpha}$ solves the equation:

$$R(\sigma_{ul,\alpha}) \;=\; \alpha\, R(\sigma_{\max}), \tag{3.4}$$

where $R(\sigma)$ is proportional to the cumulative marginal posterior distribution of the cross section of interest:

$$R(\sigma) \;\equiv\; \int_0^{\sigma} dx \int_0^1 dA\, \frac{(xAL + bL)^n}{n!}\, e^{-xAL - bL}\, \frac{e^{-\frac{1}{2}\left(\frac{A-A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\,K\,\Delta A}. \tag{3.5}$$

A straightforward calculation yields the following asymptotic expansion for $R(\sigma)$:

$$R(\sigma) \;=\; c_0 \ln\sigma \;+\; c_1 \;+\; o(1), \qquad \text{as } \sigma \to \infty, \tag{3.6}$$

with

$$\begin{aligned}
c_0 \;&\equiv\; \lim_{\sigma\to\infty} \frac{R(\sigma)}{\ln\sigma} \;=\; \frac{e^{-\frac{1}{2}\left(\frac{A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\,K\,\Delta A\,L} \sum_{i=0}^{n} \frac{(bL)^i}{i!}\, e^{-bL},\\
c_1 \;&\equiv\; \lim_{\sigma\to\infty} \left[R(\sigma) - c_0 \ln\sigma\right],\\
&=\; c_0 \int_0^{\infty} \frac{dt}{t} \left\{ \vartheta(1-t)\, e^{-\frac{1}{2}\frac{t(t-2A_0)}{\Delta A^2}} \;-\; \frac{\sum_{i=0}^{n}(bL+tL)^i/i!}{\sum_{i=0}^{n}(bL)^i/i!}\, e^{-tL} \right\}.
\end{aligned}$$

Plugging equation (3.6) into (3.4) yields:

$$\sigma_{ul,\alpha} \;=\; e^{-\frac{c_1}{c_0}(1-\alpha)}\, (\sigma_{\max})^{\alpha} \qquad \text{as } \sigma_{\max} \to \infty. \tag{3.7}$$

For the numerical values used to generate Figure 1, we find $c_0 \approx 0.002529$ and $c_1 \approx 0.5500$, so that:

$$\sigma_{ul,0.95} \;\approx\; 1.8932 \times 10^{-5}\,(\sigma_{\max})^{0.95} \qquad \text{as } \sigma_{\max} \to \infty, \tag{3.8}$$

where all cross sections are in pb. This asymptote is shown as a dashed line in the figure. It appears that the asymptotic approximation becomes reliable for $\sigma_{\max}$ values above 100 μb. For example, if we set $\sigma_{\max} = 80$ mb, the central value of the CDF measured $p\bar{p}$ total cross section at $\sqrt{s} = 1800$ GeV, then the above formula yields $\sigma_{ul,0.95} \approx 0.4$ μb, a clearly useless result. As Figure 1(b) shows, there is a wide range of $\sigma_{\max}$ values, spanning almost four orders of magnitude, over which the upper limit varies very little. Nevertheless, any choice of $\sigma_{\max}$ within that range is difficult to justify on general grounds and physicists would simply prefer to quote a result that remains finite in the limit $\sigma_{\max} \to +\infty$.

Figure 2 shows the dependence of upper limits on integrated luminosity and relative acceptance uncertainty. The top plot demonstrates that, as the integrated luminosity increases, the region dominated by prior information (small $\sigma_{\max}$ values) shortens; the influence of the data becomes stronger and the upper limits decrease accordingly. Note that in order to make this figure, the observed number of events was taken to be the most probable number of events for the given background rate and integrated luminosity in the absence of signal. This keeps the comparison between curves fair (another option would be to calculate *expected* upper limits, but this is computationally more intensive). The dashed line in the plot is actually the superposition of the three asymptotes to the solid curves. These asymptotes are almost, but not quite, identical.

The bottom plot in Figure 2 shows the dramatic effect of varying the relative acceptance uncertainty on the asymptotic behavior of the upper limits. As that uncertainty increases from 30% to 35%, the "plateau" region of cutoff values for which the upper limit is very stable almost disappears.

3.3 Information flow considerations

The information content of the posterior distribution is not only determined by *how much* prior and data information is available, but also by how much of that information is *channeled* to the parameter of interest rather than to the nuisance parameter(s). For our upper limit problem, we note that the data n cannot discriminate between σ and A since the likelihood (3.1) only depends on their product. Therefore, when we apply Bayes' theorem, the information in the data n will be shared between σ and A. How this sharing will be done, i.e. how much information goes to which parameter, depends on the prior $\pi(\sigma, A)$.

To illustrate this we return to the above example of cross section measurement, but this time we assign Gaussian priors, truncated at zero, to both the acceptance and cross section. This is not a very realistic choice, but it provides a convenient model to study how information flows from the priors to the posteriors. In particular, we wish to contrast situations where the prior information agrees with the data and situations where the two are in conflict. Assuming again $n = 5$, $b = 0.02$ pb, $A = 0.020 \pm 0.006$, and $L = 100$ pb^{-1}, we note that the cross section prior should be centered around

$\sigma_0 = (n - bL)/(AL) = 1.5$ pb for there to be no conflict between priors and data. This is illustrated in the top two entries of Table 1 for two different prior uncertainties on the cross section, 30% and 5%. In both cases the measurement affects only the uncertainties, not the mean values. For the bottom two entries of the table, the cross section prior was given a mean of 7 pb in order to demonstrate the effect of a conflict between priors and data. *Both* posteriors are now shifted with respect to the priors. Comparing measurements 3 and 4, one observes that the acceptance shift increases as the cross section is more constrained by its prior. The marginal prior and posterior densities for this example are plotted in Figures 3 and 4.

Measurement	Density	Signal Acceptance		Cross Section	
		Mean (%)	RMS/Mean (%)	Mean (pb)	RMS/Mean (%)
1:	Prior	2.0	30.0	1.5	30.0
	Posterior	2.0	27.8	1.5	27.8
2:	Prior	2.0	30.0	1.5	5.0
	Posterior	2.0	27.5	1.5	5.0
3:	Prior	2.0	30.0	7.0	30.0
	Posterior	1.4	41.2	4.9	41.2
4:	Prior	2.0	30.0	7.0	5.0
	Posterior	1.0	39.0	6.9	5.1

Table 1: Summary of four measurements of a cross section, based on five events observed over a background rate of 0.02 pb in a sample of 100 pb^{-1} of integrated luminosity. The signal acceptance and cross section priors are truncated Gaussians. Measurements 1 and 2 illustrate the effect on the posterior of a reinforcement between data and priors, whereas measurements 3 and 4 show the effect of a conflict between the two.

The main lesson from the above example is that the acceptance posterior depends not only on the data and the acceptance prior, but also on the cross section prior. This is a sensible feature if the latter is truly informative, as was for example the case at LEP with the determination of the luminosity from Bhabha events. It is a lot less sensible when one is searching for a perhaps nonexisting process like technicolor. In that case our prior beliefs about the rate of technicolor production would affect the posterior information about the effective acceptance, and hence about purely instrumental components of the latter, such as the integrated luminosity, the tracking efficiency, and the jet energy scale. Most physicists would agree that this is not a correct way to update detector parameters. The question that then arises is whether one can choose a prior such that the nuisance parameter A does not get updated by the measurement, and all the information in the data is applied to the parameter of interest σ instead.

This idea, of basing the choice of prior on considerations of information flow to the posterior, is nothing new. It leads to the so-called reference priors mentioned in section 2.1, priors that are determined by a formal rule rather than by subjective elicitation.

Such rules often involve the likelihood function of the problem at hand. Since the likelihood is proportional to the pdf, the resulting priors may suffer from a sample space dependence that violates the likelihood principle. One must therefore be careful with the interpretation of posterior distributions derived from reference priors (see [5] for a discussion). One possibility is to use these reference posteriors as "a standard to which other [posterior] distributions could be referred in order to assess the relative importance of the initial knowledge in the final results."[13]

In the next section we show how our problem of an improper posterior, equation (3.3), can indeed be avoided by more efficiently channeling the information in the data to the parameter of interest.

3.4 The correlated prior method

Normally, we think of σ and A as independent parameters and factorize the prior accordingly. However, it is also reasonable to think of the mean number of signal events $\mu_S \equiv \sigma A L$, rather than the cross section σ, as independent of A, since in practice it is μ_S and A that are independently *measured*, σ being then extracted from the ratio. This consideration is also suggested by the form of the likelihood. Indeed, whereas σ and A are *not* orthogonal with respect to expected Fisher information ($E[-\partial^2 \ln \mathcal{L}/\partial\sigma\partial A] = \sigma A L/(\sigma A + b) \neq 0$), μ_S and A are trivially so, since the likelihood is independent of A when expressed as a function of μ_S.[14] One could therefore try to factorize the prior in terms of the orthogonal parameters μ_S and A instead of σ and A:

$$\pi(\mu_S, A) \;=\; \pi(\mu_S)\,\pi(A). \tag{3.9}$$

Transforming back from (μ_S, A) to (σ, A) yields:

$$\pi(\sigma, A) \;=\; \left[\pi(\mu_S, A)\,\frac{\partial(\mu_S, A)}{\partial(\sigma, A)}\right]_{\mu_S=\sigma A L} \;=\; \Big[\pi(\mu_S)\,\pi(A)\;AL\Big]_{\mu_S=\sigma A L}. \tag{3.10}$$

We have hereby effectively introduced correlations between σ and A at the prior level. Now it turns out that with this correlated prior the information about A does not get updated at all by the measurement of n. Mathematically:

$$\begin{aligned}
p(A\,|\,n) \;&\equiv\; \int d\sigma\; p(\sigma, A\,|\,n),\\
&=\; \frac{\int d\sigma\;\mathcal{L}(\sigma, A\,|\,n)\;[\pi(\mu_S)\,\pi(A)\,AL]_{\mu_S=\sigma A L}}{\int d\sigma\int dA\;\mathcal{L}(\sigma, A\,|\,n)\;[\pi(\mu_S)\,\pi(A)\,AL]_{\mu_S=\sigma A L}},\\
&=\; \frac{\int d\mu_S\;\mathcal{L}(\mu_S\,|\,n)\,\pi(\mu_S)\,\pi(A)}{\int d\mu_S\;\mathcal{L}(\mu_S\,|\,n)\,\pi(\mu_S)\int dA\;\pi(A)},\\
&=\; \pi(A),
\end{aligned} \tag{3.11}$$

where we performed the change of variable $(\sigma, A) \to (\mu_S \equiv \sigma A L, A)$ at the third line and used the property that the likelihood only depends on σ and A through their product μ_S. The equality of $p(A\,|\,n)$ and $\pi(A)$ is a very sensible result to seek, because we are really not interested in updated information about A. We normally do not need A for other measurements, and even when we do, we simply reuse the prior $\pi(A)$. We would not want to use the posterior $p(A\,|\,n)$ since, as illustrated in section 3.3, the latter depends on the prior for σ.

With the correlated (σ, A) prior, information about A is not updated by the measurement of n, and therefore all the data information is channeled into updating σ. This turns out to solve our initial problem of a posterior density that is improper when the σ prior is improper. To see this, we start with a truncated flat prior for the number of signal events:

$$\pi(\mu_S) \;=\; \vartheta(\mu_{\max} - \mu_S)/\mu_{\max}\,, \tag{3.12}$$

so the correlated prior for (σ, A) is, from equation (3.10):

$$\pi(\sigma, A) \;=\; \frac{\vartheta(\mu_{\max} - \sigma A L)}{\mu_{\max}}\; A\,L\; \frac{e^{-\frac{1}{2}\left(\frac{A-A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\,K\,\Delta A}\,. \tag{3.13}$$

In contrast with the prior of equation (3.2), the above prior does not factorize as a function of σ and A. There are however two useful ways to characterize the prior information about σ. The first one is conditional, given a fixed value of the acceptance A:

$$\pi(\sigma\,|\,A) \;=\; \frac{\pi(\sigma, A)}{\pi(A)} \;=\; \frac{AL}{\mu_{\max}}\,\vartheta(\mu_{\max} - \sigma A L), \tag{3.14}$$

which is a simple step function that is non-zero for σ between 0 and $\mu_{\max}/(AL)$. The second one is marginal, obtained by integrating over all values of A:

$$\pi(\sigma) = \int_0^1 dA\;\pi(\sigma, A) \;=\; \frac{\operatorname{erf}\left(\frac{A_0}{\sqrt{2}\,\Delta A}\right) + \operatorname{erf}\left(\frac{u(\sigma)-A_0}{\sqrt{2}\,\Delta A}\right)}{2\mu_{\max}\,K/(A_0\,L)} \;+\; \frac{e^{-\frac{1}{2}\left(\frac{A_0}{\Delta A}\right)^2} - e^{-\frac{1}{2}\left(\frac{u(\sigma)-A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\,\mu_{\max}\,K/\Delta A}, \tag{3.15}$$

where:

$$u(\sigma) \;\equiv\; \min\left(1,\; \frac{\mu_{\max}}{\sigma L}\right).$$

This too is a step function, but with a "rounded" step (see for example Figure 5c). In the limit of small ΔA, $\pi(\sigma)$ converges to $\pi(\sigma\,|\,A_0)$.

We now turn to the calculation of the posterior density for (σ, A). For this we need the marginal data density:

$$p(n) \;=\; \int_0^{+\infty} d\sigma \int_0^1 dA\;\mathcal{L}(\sigma, A\,|\,n)\;\pi(\sigma, A) \;=\; \frac{e^{-bL}}{\mu_{\max}} \sum_{i=0}^{n} \frac{(bL)^i - e^{-\mu_{\max}}\,(\mu_{\max} + bL)^i}{i!}. \tag{3.16}$$

The posterior is then given by the ratio of $\mathcal{L}(\sigma, A\,|\,n)\;\pi(\sigma, A)$ to $p(n)$:

$$p(\sigma, A\,|\,n) \;=\; \frac{1}{K'}\;\frac{(\sigma A L + bL)^n}{n!}\;e^{-\sigma A L}\;\vartheta\left(\mu_{\max} - \sigma A L\right)\;A\,L\;\frac{e^{-\frac{1}{2}\left(\frac{A-A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\,K\,\Delta A}, \tag{3.17}$$

with:

$$K' \;\equiv\; \sum_{i=0}^{n} \frac{1}{i!}\;\left[(bL)^i - e^{-\mu_{\max}}(\mu_{\max} + bL)^i\right]. \tag{3.18}$$

By construction, the marginal acceptance posterior is equal to the marginal acceptance prior, whereas the marginal cross section posterior is:

$$p(\sigma \,|\, n) \;=\; \frac{1}{K'} \int_0^{u(\sigma)} dA \;\frac{(\sigma A L + b L)^n}{n!}\; e^{-\sigma A L}\; A L\; \frac{e^{-\frac{1}{2}\left(\frac{A-A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\, K\, \Delta A}. \tag{3.19}$$

Equations (3.13) and (3.17) are illustrated with some simple examples in Figures 5 to 10. Figure 5 shows the prior density for the case $A_0 = 0.020 \pm 0.006$, $L = 100$ pb^{-1}, and $\mu_{\max} = 100$. The corresponding posterior (setting $n = 5$ and $b = 0.02$ pb) is plotted in Figure 6. Figures 7 and 8 show what happens when the acceptance uncertainty is reduced from 0.006 to 0.002, whereas the effect of reducing the cutoff value $\mu_{\max}$ from 100 to 7 is shown in Figures 9 and 10.

The marginal posterior density for the cross section, equation (3.19), remains finite in the limit of an improper prior, as $\mu_{\max} \to +\infty$:

$$\lim_{\mu_{\max}\to\infty} p(\sigma \,|\, n) \;=\; \frac{1}{\sum_{i=0}^{n} (bL)^i / i!} \int_0^1 dA \;\frac{(\sigma A L + b L)^n}{n!}\; e^{-\sigma A L}\; A L\; \frac{e^{-\frac{1}{2}\left(\frac{A-A_0}{\Delta A}\right)^2}}{\sqrt{2\pi}\, K\, \Delta A}. \tag{3.20}$$

It is therefore possible to extract one-sided and two-sided credible intervals from the limiting posterior, which solves our problem for the single-channel case. We note that the above formula is essentially the one used by the ZEUS collaboration for their limits on selectrons and squarks, the only difference being that ZEUS included the effect of correlations between acceptance and background.[15]

The correlated prior (3.10) is generalized in Appendix A.

3.5 Example of upper limit calculations

In order to make some numerical comparisons between the correlated prior method of this note and the method described in reference [9], we introduce an effective efficiency $\epsilon \pm \Delta\epsilon$:

$$\begin{aligned} \epsilon &\equiv L\, A, \\ \Delta\epsilon &= L\, \Delta A, \end{aligned}$$

and absorb the luminosity L in the background term b. The range of ϵ is $[0, +\infty[$. With this notation, the likelihood function is:

$$\mathcal{L}(\sigma, \epsilon \,|\, n) \;=\; \frac{(\sigma\epsilon + b)^n}{n!}\; e^{-\sigma\epsilon - b}. \tag{3.21}$$

Consider a general correlated prior of the form:

$$\pi(\sigma, \epsilon) \;=\; \pi(\sigma \,|\, \epsilon)\; \pi(\epsilon) \;=\; \frac{\vartheta(\mu_{\max} - \sigma\epsilon)}{\mu_{\max}}\; \epsilon\; \pi(\epsilon). \tag{3.22}$$

A straightforward calculation then gives the cumulative marginal posterior distribution for the signal cross section σ as:

$$I\!P(\sigma \,|\, n) \;\equiv\; \int_0^{\sigma} d\sigma' \int_0^{+\infty} d\epsilon\; p(\sigma', \epsilon \,|\, n) \;=\; \int_0^{+\infty} d\epsilon\; \frac{P(n{+}1, \sigma\epsilon{+}b) - P(n{+}1, b)}{1 - P(n{+}1, b)}\; \pi(\epsilon), \tag{3.23}$$

where $P(n, x)$ is the incomplete gamma function:

$$P(n, x) \equiv \frac{1}{\Gamma(n)} \int_0^x dt\, t^{n-1} e^{-t} = \sum_{i=n}^{\infty} \frac{x^i}{i!} e^{-x} \quad \text{for integer } n > 0. \tag{3.24}$$

Upper limits are solutions of the equation $I\!P(\sigma_{ul} \,|\, n) = \alpha$ for $\alpha \in [0, 1]$. We show the results of some numerical calculations with $\alpha = 0.9$, $n = 0, \ldots, 20$, $b = 0, \ldots, 8$, and $\epsilon = 1.0$ in Tables 2 ($\Delta\epsilon = 0.0$), 3 ($\Delta\epsilon = 0.1$), and 4 ($\Delta\epsilon = 0.2$). For $\Delta\epsilon \neq 0$, two priors are used, a truncated Gaussian density:

$$\pi(\epsilon) = \frac{e^{-\frac{1}{2}\left(\frac{\epsilon - \epsilon_0}{\Delta\epsilon}\right)^2}}{\sqrt{2\pi}\, \Delta\epsilon\, \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\epsilon_0}{\sqrt{2}\, \Delta\epsilon}\right)\right]}, \tag{3.25}$$

and a gamma density:

$$\pi(\epsilon) = \frac{\epsilon^{u-1}\, e^{-\epsilon/v}}{\Gamma(u)\, v^u}, \qquad \text{with } u \equiv \left(\frac{\epsilon_0}{\Delta\epsilon}\right)^2 \text{ and } v \equiv \frac{\Delta\epsilon^2}{\epsilon_0}. \tag{3.26}$$

n	b= 0	1	2	3	4	5	6	7	8
0	2.3026	2.3026	2.3026	2.3026	2.3026	2.3026	2.3026	2.3026	2.3026
1	3.8897	3.2718	2.9949	2.8389	2.7395	2.6708	2.6206	2.5823	2.5522
2	5.3223	4.4429	3.8774	3.5228	3.2887	3.1258	3.0071	2.9172	2.8471
3	6.6808	5.7115	4.9259	4.3624	3.9666	3.6837	3.4758	3.3185	3.1964
4	7.9936	7.0000	6.0874	5.3447	4.7793	4.3570	4.0400	3.7981	3.6100
5	9.2747	8.2758	7.3058	6.4371	5.7185	5.1507	4.7096	4.3670	4.0983
6	10.5321	9.5322	8.5411	7.5993	6.7621	6.0591	5.4889	5.0337	4.6710
7	11.7709	10.7709	9.7732	8.7958	7.8798	7.0660	6.3739	5.8017	5.3352
8	12.9947	11.9947	10.9952	10.0030	9.0418	8.1478	7.3518	6.6679	6.0939
9	14.2060	13.2060	12.2061	11.2085	10.2246	9.2794	8.4040	7.6224	6.9448
10	15.4066	14.4066	13.4067	12.4073	11.4134	10.4392	9.5092	8.6497	7.8798
11	16.5981	15.5981	14.5981	13.5983	12.6004	11.6115	10.6477	9.7318	8.8859
12	17.7816	16.7816	15.7816	14.7816	13.7823	12.7867	11.8040	10.8507	9.9474
13	18.9580	17.9580	16.9580	15.9580	14.9582	13.9598	12.9674	11.9916	11.0484
14	20.1280	19.1280	18.1280	17.1280	16.1281	15.1286	14.1318	13.1434	12.1746
15	21.2924	20.2924	19.2924	18.2924	17.2924	16.2926	15.2938	14.2990	13.3151
16	22.4516	21.4516	20.4516	19.4516	18.4516	17.4516	16.4521	15.4543	14.4621
17	23.6061	22.6061	21.6061	20.6061	19.6061	18.6061	17.6063	16.6071	15.6107
18	24.7563	23.7563	22.7563	21.7563	20.7563	19.7563	18.7563	17.7567	16.7582
19	25.9025	24.9025	23.9025	22.9025	21.9025	20.9025	19.9025	18.9027	17.9033
20	27.0451	26.0451	25.0451	24.0451	23.0451	22.0451	21.0451	20.0451	19.0454

Table 2: Upper limits at the 90% credibility level on the cross section σ of a Poisson process with effective efficiency $\epsilon = 1.0$, as a function of the total number n of events observed and the expected number b of background events. There is no uncertainty on the value of ϵ. In this table and the following two, the units of σ and ϵ are arbitrary but consistent with the product $\sigma \cdot \epsilon$ being unitless.

n	b= 0	1	2	3	4	5	6	7	8
0	2.3296	2.3296	2.3296	2.3296	2.3296	2.3296	2.3296	2.3296	2.3296
1	3.9473	3.3137	3.0315	2.8730	2.7721	2.7024	2.6515	2.6128	2.5823
2	5.4130	4.5067	3.9280	3.5667	3.3288	3.1634	3.0430	2.9519	2.8809
3	6.8069	5.8038	4.9961	4.4199	4.0166	3.7290	3.5179	3.3584	3.2345
4	8.1573	7.1254	6.1832	5.4206	4.8427	4.4124	4.0901	3.8444	3.6535
5	9.4778	8.4374	7.4321	6.5367	5.7996	5.2194	4.7700	4.4215	4.1486
6	10.7765	9.7324	8.7018	7.7274	6.8655	6.1448	5.5624	5.0985	4.7296
7	12.0582	11.0115	9.9713	8.9566	8.0101	7.1730	6.4639	5.8795	5.4042
8	13.3264	12.2775	11.2329	10.1998	9.2033	8.2805	7.4624	6.7620	6.1758
9	14.5835	13.5325	12.4852	11.4439	10.4208	9.4420	8.5393	7.7364	7.0427
10	15.8313	14.7784	13.7288	12.6834	11.6471	10.6353	9.6733	8.7876	7.9971
11	17.0713	16.0165	14.9649	13.9168	12.8740	11.8442	10.8442	9.8974	9.0264
12	18.3044	17.2479	16.1945	15.1441	14.0976	13.0583	12.0360	11.0477	10.1148
13	19.5316	18.4735	17.4183	16.3660	15.3169	14.2725	13.2375	12.2232	11.2463
14	20.7536	19.6939	18.6370	17.5829	16.5318	15.4841	14.4422	13.4123	12.4062
15	21.9709	20.9098	19.8513	18.7955	17.7425	16.6925	15.6464	14.6076	13.5832
16	23.1841	22.1216	21.0616	20.0042	18.9495	17.8975	16.8486	15.8045	14.7691
17	24.3936	23.3297	22.2683	21.2094	20.1530	19.0992	18.0482	17.0007	15.9587
18	25.5997	24.5345	23.4717	22.4113	21.3534	20.2979	19.2451	18.1951	17.1490
19	26.8028	25.7364	24.6722	23.6104	22.5510	21.4939	20.4394	19.3874	18.3385
20	28.0030	26.9354	25.8700	24.8069	23.7460	22.6875	21.6313	20.5775	19.5265

0	2.3293	2.3293	2.3293	2.3293	2.3293	2.3293	2.3293	2.3293	2.3293
1	3.9464	3.3131	3.0310	2.8726	2.7717	2.7020	2.6512	2.6124	2.5820
2	5.4114	4.5058	3.9274	3.5661	3.3283	3.1629	3.0425	2.9515	2.8805
3	6.8045	5.8023	4.9951	4.4192	4.0160	3.7285	3.5174	3.3579	3.2341
4	8.1540	7.1232	6.1817	5.4196	4.8419	4.4118	4.0894	3.8439	3.6530
5	9.4737	8.4345	7.4300	6.5352	5.7985	5.2185	4.7692	4.4208	4.1480
6	10.7713	9.7285	8.6990	7.7254	6.8640	6.1437	5.5615	5.0977	4.7289
7	12.0520	11.0067	9.9677	8.9540	8.0082	7.1715	6.4627	5.8785	5.4033
8	13.3191	12.2717	11.2283	10.1964	9.2007	8.2786	7.4609	6.7608	6.1748
9	14.5751	13.5256	12.4797	11.4396	10.4175	9.4395	8.5374	7.7349	7.0414
10	15.8217	14.7704	13.7223	12.6781	11.6430	10.6321	9.6708	8.7857	7.9956
11	17.0605	16.0074	14.9573	13.9105	12.8689	11.8402	10.8411	9.8950	9.0244
12	18.2923	17.2376	16.1857	15.1368	14.0916	13.0535	12.0321	11.0447	10.1124
13	19.5182	18.4619	17.4083	16.3576	15.3099	14.2667	13.2329	12.2194	11.2433
14	20.7388	19.6810	18.6259	17.5734	16.5237	15.4774	14.4366	13.4078	12.4026
15	21.9548	20.8956	19.8389	18.7848	17.7333	16.6847	15.6399	14.6022	13.5788
16	23.1666	22.1060	21.0478	19.9922	18.9390	17.8885	16.8410	15.7981	14.7639
17	24.3746	23.3127	22.2532	21.1960	20.1413	19.0891	18.0395	16.9933	15.9525
18	25.5792	24.5161	23.4553	22.3967	21.3405	20.2866	19.2352	18.1866	17.1418
19	26.7808	25.7165	24.6543	23.5944	22.5367	21.4814	20.4284	19.3778	18.3303
20	27.9795	26.9140	25.8507	24.7895	23.7304	22.6736	21.6190	20.5668	19.5171

Table 3: Upper limits at the 90% credibility level on the cross section of a Poisson process with effective efficiency $\epsilon = 1.0 \pm 0.1$, as a function of the total number n of events observed and the expected number b of background events. Top: the ϵ prior is a truncated Gaussian distribution. Bottom: the ϵ prior is a gamma distribution.

n	b= 0	1	2	3	4	5	6	7	8
0	2.4197	2.4197	2.4197	2.4197	2.4197	2.4197	2.4197	2.4197	2.4197
1	4.1389	3.4524	3.1529	2.9862	2.8804	2.8076	2.7546	2.7142	2.6824
2	5.7142	4.7175	4.0954	3.7120	3.4615	3.2881	3.1622	3.0671	2.9930
3	7.2244	6.1082	5.2276	4.6099	4.1821	3.8791	3.6576	3.4906	3.3612
4	8.6968	7.5383	6.4979	5.6705	5.0519	4.5956	4.2558	3.9979	3.7980
5	10.1444	8.9683	7.8467	6.8634	6.0660	5.4456	4.9691	4.6017	4.3150
6	11.5744	10.3877	9.2285	8.1471	7.2043	6.4264	5.8042	5.3123	4.9232
7	12.9914	11.7963	10.6188	9.4827	8.4366	7.5238	6.7596	6.1356	5.6316
8	14.3984	13.1959	12.0073	10.8424	9.7310	8.7147	7.8247	7.0708	6.4453
9	15.7973	14.5883	13.3914	12.2104	11.0608	9.9730	8.9817	8.1099	7.3640
10	17.1899	15.9750	14.7710	13.5796	12.4079	11.2746	10.2087	9.2383	8.3813
11	18.5771	17.3569	16.1466	14.9473	13.7621	12.6010	11.4842	10.4379	9.4854
12	19.9599	18.7348	17.5188	16.3127	15.1181	13.9399	12.7903	11.6895	10.6608
13	21.3390	20.1094	18.8882	17.6761	16.4741	15.2845	14.1140	12.9762	11.8906
14	22.7148	21.4811	20.2551	19.0376	17.8293	16.6313	15.4471	14.2850	13.1590
15	24.0879	22.8503	21.6200	20.3975	19.1835	17.9786	16.7848	15.6065	14.4531
16	25.4586	24.2175	22.9831	21.7560	20.5367	19.3259	18.1245	16.9351	15.7631
17	26.8272	25.5827	24.3446	23.1133	21.8892	20.6730	19.4653	18.2674	17.0826
18	28.1939	26.9463	25.7047	24.4695	23.2410	22.0199	20.8065	19.6018	18.4076
19	29.5590	28.3085	27.0636	25.8247	24.5922	23.3665	22.1480	20.9375	19.7358
20	30.9226	29.6694	28.4215	27.1791	25.9428	24.7128	23.4897	22.2739	21.0661

0	2.4119	2.4119	2.4119	2.4119	2.4119	2.4119	2.4119	2.4119	2.4119
1	4.1216	3.4408	3.1428	2.9767	2.8713	2.7987	2.7458	2.7055	2.6739
2	5.6863	4.7001	4.0819	3.7002	3.4506	3.2778	3.1522	3.0574	2.9835
3	7.1852	6.0824	5.2091	4.5948	4.1689	3.8669	3.6462	3.4797	3.3507
4	8.6457	7.5023	6.4724	5.6509	5.0356	4.5812	4.2425	3.9855	3.7862
5	10.0810	8.9210	7.8122	6.8377	6.0455	5.4282	4.9535	4.5874	4.3017
6	11.4983	10.3282	9.1836	8.1134	7.1782	6.4050	5.7857	5.2957	4.9080
7	12.9022	11.7241	10.5623	9.4394	8.4032	7.4971	6.7373	6.1162	5.6141
8	14.2957	13.1107	11.9385	10.7881	9.6887	8.6814	7.7975	7.0477	6.4250
9	15.6810	14.4898	13.3097	12.1443	11.0082	9.9314	8.9482	8.0821	7.3401
10	17.0596	15.8628	14.6760	13.5009	12.3440	11.2233	10.1674	9.2046	8.3529
11	18.4327	17.2307	16.0380	14.8555	13.6859	12.5388	11.4338	10.3968	9.4514
12	19.8012	18.5945	17.3964	16.2075	15.0291	13.8659	12.7295	11.6398	10.6199
13	21.1657	19.9547	18.7516	17.5570	16.3717	15.1979	14.0418	12.9166	11.8414
14	22.5269	21.3118	20.1042	18.9044	17.7132	16.5315	15.3626	14.2143	13.1003
15	23.8851	22.6663	21.4545	20.2500	19.0534	17.8653	16.6874	15.5239	14.3837
16	25.2408	24.0185	22.8028	21.5939	20.3924	19.1987	18.0138	16.8398	15.6821
17	26.5943	25.3688	24.1493	22.9364	21.7304	20.5316	19.3407	18.1590	16.9892
18	27.9457	26.7172	25.4944	24.2777	23.0674	21.8640	20.6679	19.4798	18.3013
19	29.2955	28.0641	26.8381	25.6178	24.4036	23.1959	21.9950	20.8014	19.6160
20	30.6436	29.4095	28.1806	26.9570	25.7392	24.5274	23.3221	22.1235	20.9324

Table 4: Upper limits at the 90% credibility level on the cross section of a Poisson process with effective efficiency $\epsilon = 1.0 \pm 0.2$, as a function of the total number n of events observed and the expected number b of background events. Top: the ϵ prior is a truncated Gaussian distribution. Bottom: the ϵ prior is a gamma distribution.

One observes that upper limits obtained with the gamma prior are slightly lower than those obtained with the truncated Gaussian prior. This can be understood from equation (3.23), whose right-hand side is the integral of the prior density $\pi(\epsilon)$ times a rising step function. Since the Gaussian density has a heavier lower tail than the gamma density, it suffers a bigger loss at the cutting edge of the step. However, that edge moves left as σ increases, reducing its cutting effect. Therefore, a given value α of the integral (3.23) is attained with a larger upper limit σ when $\pi(\epsilon)$ is Gaussian than when it is gamma.

Comparing the bottom section of Table 3 with Table 1 in reference [9], one notes that upper limits from the correlated prior method are lower. This is to be expected, since the correlated prior method channels all the data information into the parameter of interest. In contrast, the method of [9] updates the nuisance prior non-trivially, thereby limiting the amount of information flowing from the data to the parameter of interest.

Software to calculate upper limits as described in this section is available from the author.

3.6 Effect of background uncertainties

In this section we show that the situation with background uncertainties is radically different from that with acceptance uncertainties. Consider again the likelihood of equation (3.1), but this time assume that, a priori, the acceptance is exactly known, whereas the background rate b and the signal cross section σ are only known to be bounded by upper limits b_M and σ_M respectively:

$$\pi(\sigma, b) \;=\; \frac{\vartheta(\sigma_M - \sigma)}{\sigma_M}\,\frac{\vartheta(b_M - b)}{b_M}. \tag{3.27}$$

The marginal density of n is:

$$\begin{aligned} p(n) \;&=\; \int_0^{\infty} d\sigma \int_0^{\infty} db\, \frac{(\sigma AL + bL)^n}{n!}\, e^{-\sigma AL - bL}\, \frac{\vartheta(\sigma_M - \sigma)}{\sigma_M}\, \frac{\vartheta(b_M - b)}{b_M}, \\ &=\; \frac{1}{A\,\sigma_M\, b_M\, L^2}\Bigg\{ n + 1 - \sum_{j=0}^{n} \frac{n+1-j}{j!}\Big[(b_M L)^j e^{-b_M L} + (\sigma_M AL)^j e^{-\sigma_M AL} \\ &\qquad - (\sigma_M AL + b_M L)^j e^{-\sigma_M AL - b_M L}\Big]\Bigg\}. \end{aligned} \tag{3.28}$$

The posterior density $p(\sigma, b\,|\,n)$ remains well defined in the limit $\sigma_M, b_M \to +\infty$, when the prior becomes improper:

$$\lim_{\substack{\sigma_M\to\infty\\ b_M\to\infty}} p(\sigma, b\,|\,n) \;=\; \lim_{\substack{\sigma_M\to\infty\\ b_M\to\infty}} \frac{\mathcal{L}(\sigma, b\,|\,n)\;\pi(\sigma, b)}{p(n)} \;=\; \frac{A\,L^2}{n+1}\,\frac{(\sigma AL + bL)^n}{n!}\, e^{-\sigma AL - bL}. \tag{3.29}$$

The limiting marginal posterior for σ is:

$$\lim_{\substack{\sigma_M\to\infty\\ b_M\to\infty}} p(\sigma\,|\,n) \;=\; \lim_{\substack{\sigma_M\to\infty\\ b_M\to\infty}} \int_0^{\infty} db\; p(\sigma, b\,|\,n) \;=\; \frac{A\,L}{n+1}\, e^{-\sigma AL} \sum_{i=0}^{n} \frac{(\sigma AL)^i}{i!}. \tag{3.30}$$

In contrast with the situation for acceptance uncertainties, here we find that there is always enough information in the data to make the posterior a proper density, even when both the signal and background priors are improper.

This is a good place to point out the difference between the effects of flat improper priors for signal and background. As explained in footnote 2, a flat signal prior from zero to infinity yields conservative upper limits when compared to a flat signal prior from zero to some finite boundary. However, a flat background prior from zero to infinity has the opposite effect, since it expresses the belief that the true background must be quite large and will therefore cause a larger background subtraction. This effect is illustrated in Figure 11.

3.7 Expected upper limits

It is often useful to look at *expected* upper limits, as a means of quantifying the sensitivity of one's measurement, or for the purpose of designing a new experiment. One must then decide over what ensemble expectations are to be evaluated. If for n observed events we denote the α-credibility level upper limit on the cross section by $\sigma_{ul}^{\alpha}(n)$, then the expected upper limit can be written very generally as:

$$\mathrm{E}\big[\sigma_{ul}^{\alpha}(N)\big] \;=\; \sum_{n=0}^{\infty} w_n \, \sigma_{ul}^{\alpha}(n), \tag{3.31}$$

where the weights w_n sum to 1. An obvious choice is to use the marginal data density of equation (3.16) and set $w_n \equiv p(n)$. This will work as long as $\mu_{\max}$ is finite, but unfortunately $p(n)$ vanishes in the limit $\mu_{\max} \to +\infty$.

Another interesting choice is to calculate expected upper limits under the assumption of zero signal:

$$w_n \;=\; \frac{(bL)^n}{n!}\, e^{-bL}. \tag{3.32}$$

If there is an uncertainty Δb on the background rate b, one can use a Gaussian prior for b and set:

$$w_n \;=\; c \int_0^{+\infty} db\, \frac{(bL)^n}{n!}\, e^{-bL}\, \frac{e^{-\frac{1}{2}\left(\frac{b-b_0}{\Delta b}\right)^2}}{\sqrt{2\pi}\,\Delta b}, \tag{3.33}$$

where c is a normalization constant chosen to make the w_n sum to 1. Finally, if nothing is known about the background and one chooses to represent this ignorance by a flat improper prior, then it is no longer possible to calculate expected upper limits, since the corresponding marginal data density (3.28) vanishes in the limit $b_M \to \infty$.[3] In other words, for a very large prior cutoff b_M on the background, n can be very large, and hence also the upper limit; the expected upper limit then diverges as the cutoff goes to infinity.

[3] Although it is not possible to calculate *expected* upper limits as $b_M \to \infty$, once data are observed it is still possible to extract a *measured* upper limit, as shown in section 3.6. This should not surprise, as expected limits use only prior information, whereas measured limits use both prior and data information.

4 Upper limits in multiple-channel searches

For searches involving more than one channel there is a variety of situations to consider, depending on how uncertainties in different channels are correlated. We examine a few commonly occurring examples in the following subsections.

4.1 Correlated uncertainties on signal efficiency

Suppose that one is fitting a histogram with common signal efficiency and signal cross section in all bins. The data is then a vector of event counts $\vec{n} \equiv (n_i)$ and the likelihood has the form:

$$\mathcal{L}(\sigma, A \,|\, \vec{n}) \;=\; \prod_{i=1}^{N} \frac{(\sigma A L t_{si} + b L t_{bi})^{n_i}}{n_i!}\; e^{-\sigma A L t_{si} - b L t_{bi}}, \tag{4.1}$$

where t_{si} and t_{bi} are normalized signal and background templates respectively, and the product is over histogram bins. This situation is not really different from the single-channel case, in that a truncated Gaussian prior for A and a flat prior for σ will yield a divergent predictive density. Here too a solution is to factorize the prior in terms of the expected total number of signal events $\mu_S \equiv \sigma A L$ and the acceptance A.

4.2 Correlated uncertainties on template shapes

Continuing with the example of a histogram fit, suppose now that in each bin i the signal template t_{si} is only known within $\pm\Delta t_{si}$, and that this uncertainty is correlated across all bins. The likelihood can then be rewritten as:

$$\mathcal{L}(\sigma, A, u \,|\, \vec{n}) \;=\; \prod_{i=1}^{N} \frac{[\sigma A L(t_{si} + u\Delta t_{si}) + b L t_{bi}]^{n_i}}{n_i!}\; e^{-\sigma A L(t_{si} + u\Delta t_{si}) - b L t_{bi}}, \tag{4.2}$$

where u is an additional parameter for which an appropriate prior must be chosen. A common choice is a truncated Gaussian; the resulting predictive density will then be finite as long as the lower boundary $u_{\rm low}$ of the integration region for u satisfies the constraint:

$$u_{\rm low} \;>\; -\frac{\sum_{i=1}^{N} t_{si}}{\sum_{i=1}^{N} \Delta t_{si}}. \tag{4.3}$$

For physical reasons one usually wants the quantity $t_{si} + u\Delta t_{si}$ to be positive in all bins. This is sufficient for the above constraint to be satisfied, provided not all the $t_{si}/\Delta t_{si}$ are equal.

Since divergences can easily be avoided by carefully truncating the prior for u, the correlated prior method is not really needed here. Nevertheless, in order to be able to make meaningful comparisons between the single and multiple-channel cases, it is useful to generalize that method. To this end, we note that the total expected number of signal events is given by:

$$\mu_S \;=\; \sigma\, A_{\rm tot}\, L, \tag{4.4}$$

where:

$$A_{\rm tot} \;=\; A\,\sum_{i=1}^{N}(t_{si} + u\Delta t_{si}) \tag{4.5}$$

is the "total effective signal acceptance" for all channels combined. The correlated prior of equation (3.10) then takes the form:

$$\pi(\sigma, A_{\rm tot}) \;=\; \bigl[\pi(\mu_S)\;A_{\rm tot}\;L\;\pi(A_{\rm tot})\bigr]_{\mu_S=\sigma A_{\rm tot}L}. \tag{4.6}$$

However, this form is inadequate since we have separate prior information about A and u, and not just about $A_{\rm tot}$. Accordingly, we set:

$$\pi(\sigma, A, u) \;=\; \bigl[\pi(\mu_S)\;A_{\rm tot}\;L\;\pi(A)\;\pi(u)\bigr]_{\mu_S=\sigma A_{\rm tot}L}. \tag{4.7}$$

Note that because of this modified prior factorization, one of the properties of the single-channel correlated prior, namely that it leads to an acceptance posterior equal to its prior (equation 3.11), is no longer exactly realized.

4.3 Uncorrelated uncertainties

An example here is combining measurements in several independent channels, so the likelihood is:

$$\mathcal{L}(\sigma, \vec{A}\,|\,\vec{n}) \;=\; \prod_{i=1}^{N_C}\frac{(\sigma A_i L t_{si} + bLt_{bi})^{n_i}}{n_i!}\;e^{-\sigma A_i L t_{si} - bLt_{bi}}, \tag{4.8}$$

where the product is now over N_C channels, the acceptances A_i are independent, and there is proper prior information about each A_i. As shown in Appendix B, this multidimensionality of $\vec{A}$ helps, in the sense that there is enough information in the data and priors to make the cross section posterior proper regardless of its prior.

Here too however, we may be interested in comparing single-channel upper limits with the combined upper limit, so that a generalization of the correlated prior method would be useful. This can again be done by introducing the total expected number of signal events of equation (4.4), the "total effective signal acceptance" now being given by:

$$A_{\rm tot} \;=\; \sum_{i=1}^{N_C} A_i\,t_{si}. \tag{4.9}$$

The generalized correlated prior is then:

$$\pi(\sigma, \vec{A}) \;=\; \left[\pi(\mu_S)\;A_{\rm tot}\;L\;\prod_{i=1}^{N_C}\pi(A_i)\right]_{\mu_S=\sigma A_{\rm tot}L}. \tag{4.10}$$

With a flat signal rate prior and truncated Gaussian acceptance prior, this yields:

$$\pi(\sigma, \vec{A}) \;=\; \frac{\vartheta(\mu_{\rm max} - \sigma A_{\rm tot}L)}{\mu_{\rm max}}\;A_{\rm tot}\;L\;\prod_{i=1}^{N_C}\frac{e^{-\frac{1}{2}\left(\frac{A_i - A_{0i}}{\Delta A_i}\right)^2}}{\sqrt{2\pi}\,K_i\,\Delta A_i}, \tag{4.11}$$

$$\text{with}\;\; K_i \;\equiv\; \frac{1}{2}\Bigl[1 + {\rm erf}\Bigl(\frac{A_{0i}}{\sqrt{2}\,\Delta A_i}\Bigr)\Bigr].$$

4.4 General case

In general the likelihood is a product over N_C channels, and within each channel i it is a product over N_i histogram bins:

$$\mathcal{L}(\sigma, \vec{A} \,|\, \vec{n}) \;=\; \prod_{i=1}^{N_C} \prod_{j=1}^{N_i} \frac{[\sigma A_i L(t_{sij} + u\Delta t_{sij}) + b_i L t_{bij}]^{n_{ij}}}{n_{ij}!} \; e^{-\sigma A_i L(t_{sij}+u\Delta t_{sij}) - b_i L t_{bij}}. \quad (4.12)$$

The total effective signal acceptance is now:

$$A_{\rm tot} \;=\; \sum_{i=1}^{N_C} \sum_{j=1}^{N_i} A_i\,(t_{sij} + u\Delta t_{sij}), \quad (4.13)$$

and the appropriate generalization of the correlated prior is given by:

$$\pi(\sigma, \vec{A}, u) \;=\; \left[\pi(\mu_S)\; A_{\rm tot}\; L\; \prod_{i=1}^{N_C} \pi(A_i)\; \pi(u)\right]_{\mu_S = \sigma A_{\rm tot} L}. \quad (4.14)$$

Further generalizations along the same lines are of course possible.

5 Computational aspects

When working with several uncertainties, the marginal posterior density for the cross section σ is proportional to a multidimensional integral over nuisance parameters:

$$p(\sigma \,|\, \vec{n}) \;=\; \int\! d\vec{A} \int\! d\vec{b}\; p(\sigma, \vec{A}, \vec{b} \,|\, \vec{n}) \;\propto\; \int\! d\vec{A} \int\! d\vec{b}\; \mathcal{L}(\sigma, \vec{A}, \vec{b} \,|\, \vec{n})\; \pi(\sigma, \vec{A}, \vec{b}), \quad (5.1)$$

where the list of nuisance parameters includes a vector of acceptances $\vec{A}$ and a vector of backgrounds $\vec{b}$; the integrals are over the whole $(\vec{A}, \vec{b})$ parameter space. As shown in section 3.6, background uncertainties do not cause problems, essentially because the corresponding nuisance parameters are additive. The background component of the prior can therefore be factored out:

$$\pi(\sigma, \vec{A}, \vec{b}) \;=\; \pi(\sigma, \vec{A})\; \pi(\vec{b}), \quad (5.2)$$

whereas the $(\sigma, \vec{A})$ component should have a correlated form as discussed in the previous section. In order to extract an upper limit σ_{ul} at the α credibility level, we need to integrate $p(\sigma \,|\, \vec{n})$ with respect to σ:

$$\int_0^{\sigma_{ul}} d\sigma\; p(\sigma \,|\, \vec{n}) \;=\; \alpha. \quad (5.3)$$

In the following we will find it convenient to work with the *un*normalized posterior density

$$q(\sigma, \vec{A}, \vec{b} \,|\, \vec{n}) \;\equiv\; \mathcal{L}(\sigma, \vec{A}, \vec{b} \,|\, \vec{n})\; \pi(\sigma, \vec{A}, \vec{b}) \quad (5.4)$$

instead of $p(\sigma, \vec{A}, \vec{b} \,|\, \vec{n})$ itself. Writing the marginal data density as

$$m(\vec{n}) \;\equiv\; \int_0^\infty d\sigma \int d\vec{A} \int d\vec{b}\; q(\sigma, \vec{A}, \vec{b} \,|\, \vec{n}), \tag{5.5}$$

the defining equation for σ_{ul} becomes:

$$\frac{1}{m(\vec{n})} \int_0^{\sigma_{ul}} d\sigma \int d\vec{A} \int d\vec{b}\; q(\sigma, \vec{A}, \vec{b} \,|\, \vec{n}) \;=\; \alpha. \tag{5.6}$$

The easiest and most efficient way to calculate the integrals in this equation is by a Monte Carlo importance sampling algorithm. Let $g(\sigma, \vec{A}, \vec{b})$ be a density that is easy to sample from; the strong law of large numbers then yields:

$$\int d\sigma \int d\vec{A} \int d\vec{b}\; q(\sigma, \vec{A}, \vec{b} \,|\, \vec{n}) \;=\; \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{q(\sigma_i, \vec{A}_i, \vec{b}_i \,|\, \vec{n})}{g(\sigma_i, \vec{A}_i, \vec{b}_i)} \tag{5.7}$$

provided the $(\sigma_i, \vec{A}_i, \vec{b}_i)$ *are sampled from* $g(\sigma, \vec{A}, \vec{b})$. In this equation we deliberately left the σ integration limits unspecified. To find σ_{ul} we will in fact need to calculate the σ integral over a fine grid of upper integration limits. This can be done by executing the following steps a large number of times:

1. Generate $(\sigma, \vec{A}, \vec{b})$ according to g;
2. Compute the weight: $w(\sigma, \vec{A}, \vec{b}) \;\equiv\; q(\sigma, \vec{A}, \vec{b} \,|\, \vec{n}) \;/\; g(\sigma, \vec{A}, \vec{b})$; (5.8)
3. Obtain $q(\sigma \,|\, \vec{n}) \equiv \int d\vec{A} \int d\vec{b}\; q(\sigma, \vec{A}, \vec{b} \,|\, \vec{n})$ by histogramming σ with the weight from step 2.

Once the σ histogram has been generated, it can be converted into a properly normalized cumulative distribution from which σ_{ul} can be estimated.

5.1 Choice of importance sampling density

The question of how to choose the importance sampling density g can be addressed by studying the variance of the α credibility upper limit σ_{ul}. According to the central limit theorem (see Appendix C), this variance is asymptotically given by:

$$\mathrm{Var}(\sigma_{ul}) \;=\; \frac{\displaystyle\int_0^\infty d\sigma \int d\vec{A} \int d\vec{b}\; \Big\{\big[\alpha - \vartheta(\sigma_{ul} - \sigma)\big] q(\sigma, \vec{A}, \vec{b} \,|\, \vec{n})\Big\}^2 \Big/ g(\sigma, \vec{A}, \vec{b})}{N\; \left[\displaystyle\int d\vec{A} \int d\vec{b}\; q(\sigma_{ul}, \vec{A}, \vec{b} \,|\, \vec{n})\right]^2}, \tag{5.9}$$

and is minimized for $g = g_{\rm opt}$, where:

$$\begin{aligned} g_{\rm opt}(\sigma, \vec{A}, \vec{b}) \;&=\; \frac{\big|\alpha - \vartheta(\sigma_{ul} - \sigma)\big|}{2\;\alpha\;(1-\alpha)}\; \frac{1}{m(\vec{n})}\; q(\sigma, \vec{A}, \vec{b} \,|\, \vec{n}), \\ &=\; \left[\frac{1}{2\;\alpha}\;\vartheta(\sigma_{ul} - \sigma) \;+\; \frac{1}{2\;(1-\alpha)}\;\vartheta(\sigma - \sigma_{ul})\right] \frac{1}{m(\vec{n})}\; q(\sigma, \vec{A}, \vec{b} \,|\, \vec{n}). \end{aligned} \tag{5.10}$$

The optimal importance sampling density is therefore identical to the posterior density except for a difference in normalization above and below the upper limit σ_{ul}. This difference is such that σ_{ul}, the $\alpha^{\rm th}$ quantile of the posterior, is in fact the median, or $50^{\rm th}$ percentile, of $g_{\rm opt}$:

$$\int_0^{\sigma_{ul}} d\sigma \int d\vec{A} \int d\vec{b}\; g_{\rm opt}(\sigma, \vec{A}, \vec{b}) \;=\; \int_{\sigma_{ul}}^{+\infty} d\sigma \int d\vec{A} \int d\vec{b}\; g_{\rm opt}(\sigma, \vec{A}, \vec{b}). \tag{5.11}$$

Knowledge of the optimal importance sampling density is of limited use since we do not know σ_{ul} a priori, and since it is generally not easy to generate random numbers from the posterior density itself. It can serve as a guide however, suggesting that if a preliminary estimate of σ_{ul} is available, the accuracy of the estimate can be improved more efficiently by sampling equally below and above its preliminary value.

Generic recommendations on importance sampling are motivated and discussed in reference [16]. Undoubtedly the most important recommendation is that the tails of the importance sampling density should not decay more quickly than the tails of the function one is integrating (the posterior density). If this condition is violated, the Monte Carlo algorithm will occasionally generate large weights, resulting in large fluctuations in the estimate of the integral of interest. In general the prior, when proper, can safely be used as importance sampling density. This is of direct applicability to nuisance parameters, since the latter tend to have proper, informative priors in most high-energy physics problems. On the other hand, interest parameters are usually given some kind of non-informative, improper prior, in which case it is very inefficient, and maybe even impossible to sample from the prior. A solution is to use a distribution from the normal or t family instead, and to adjust its shape (making sure its tail does not decay too quickly) by checking the marginal posterior density for the parameter of interest after an initial run of the program.

5.2 Calculation of uncertainties

The uncertainty on σ_{ul} can be approximated by the square root of the asymptotic variance formula (5.9). To implement this in an importance sampling Monte Carlo calculation requires that one keep track of sums of squared weights in addition to sums of weights. Assume that we have at our disposal a total of N Monte Carlo samples $(\sigma_i, \vec{A}_i, \vec{b}_i)$, $i = 1, \ldots, N$, drawn from an importance sampling density g. Each sample carries a weight $w(\sigma_i, \vec{A}_i, \vec{b}_i)$ given by equation (5.8). We introduce the notation:

$$S_N(\xi) \equiv \sum_{\sigma_i \leq \xi} w(\sigma_i, \vec{A}_i, \vec{b}_i) \qquad\qquad S_N^\star \equiv \sum_{i=1}^{N} w(\sigma_i, \vec{A}_i, \vec{b}_i) \tag{5.12}$$

$$T_N(\xi) \equiv \sum_{\sigma_i \leq \xi} \Big[w(\sigma_i, \vec{A}_i, \vec{b}_i)\Big]^2 \qquad\qquad T_N^\star \equiv \sum_{i=1}^{N} \Big[w(\sigma_i, \vec{A}_i, \vec{b}_i)\Big]^2 \tag{5.13}$$

In the large N limit, these sums are related to the following integrals:

$$\int_0^{\xi} d\sigma \int d\vec{A} \int d\vec{b}\; q(\sigma, \vec{A}, \vec{b} \,|\, \vec{n}) = \int_{\sigma\leq\xi} \frac{q(\sigma, \vec{A}, \vec{b} \,|\, \vec{n})}{g(\sigma, \vec{A}, \vec{b})}\; dG(\sigma, \vec{A}, \vec{b}) = \frac{S_N(\xi)}{N},$$

$$\int_0^{\xi} d\sigma \int d\vec{A} \int d\vec{b}\; \frac{\left[q(\sigma, \vec{A}, \vec{b} \,|\, \vec{n})\right]^2}{g(\sigma, \vec{A}, \vec{b})} = \int_{\sigma\leq\xi} \left[\frac{q(\sigma, \vec{A}, \vec{b} \,|\, \vec{n})}{g(\sigma, \vec{A}, \vec{b})}\right]^2 dG(\sigma, \vec{A}, \vec{b}) = \frac{T_N(\xi)}{N}.$$

With these formulae, equation (5.9) can be rewritten in terms of the weight sums:

$$\mathrm{Var}(\sigma_{ul}) \;=\; \left[\frac{\sqrt{\alpha^2\, T_N^{\star} + (1-2\alpha)\, T_N(\sigma_{ul})}}{S_N^{\star}\; p(\sigma_{ul} \,|\, \vec{n})}\right]^2, \tag{5.14}$$

where $p(\sigma_{ul} \,|\, \vec{n})$, the marginal posterior density evaluated at the upper limit, can be estimated as follows:

$$p(\sigma_{ul} \,|\, \vec{n}) \;=\; \int d\vec{A} \int d\vec{b}\; \frac{q(\sigma_{ul}, \vec{A}, \vec{b} \,|\, \vec{n})}{m(\vec{n})} \;\approx\; \frac{S_N(\sigma_{ul} + \Delta\sigma) - S_N(\sigma_{ul})}{S_N^{\star}\; \Delta\sigma}, \tag{5.15}$$

with $\Delta\sigma$ a suitable bin width.

It is often useful to have a measure of the adequacy of the importance sampling density g for the calculation of interest. This can be obtained by considering the variance that would result if the importance sampling density were the posterior density itself. In this case, all the weights would be equal: $w(\sigma, \vec{A}, \vec{b}) = m(\vec{n})$, so that:

$$\mathrm{Var}(\sigma_{ul})\Big|_{g=p(\cdot\,|\,\vec{n})} \;=\; \frac{\alpha\,(1-\alpha)}{N}\; \frac{1}{\left[p(\sigma_{ul} \,|\, \vec{n})\right]^2}\,. \tag{5.16}$$

Note that this is not the smallest variance that can be achieved, since we have seen that the optimal importance sampling density for quantile estimation differs from the posterior density by a discontinuity at the quantile of interest. Nevertheless, if our purpose was to write a *general* program that could be used to calculate any arbitrary characteristic of the posterior density, then the optimal choice of importance sampling density would indeed be the posterior itself. The ratio of the above variance to the one of equation (5.14) is called the relative numerical efficiency (RNE)[16]:

$$\mathrm{RNE} \;\equiv\; \frac{\mathrm{Var}(\sigma_{ul})\big|_{g=p(\cdot\,|\,\vec{n})}}{\mathrm{Var}(\sigma_{ul})} \;=\; \frac{\alpha\,(1-\alpha)}{\alpha^2 T_N^{\star} \;+\; (1-2\alpha)\, T_N(\sigma_{ul})}\; \frac{(S_N^{\star})^2}{N}. \tag{5.17}$$

Small RNE values indicate that there exists an importance sampling density, namely the posterior, that does not have to be tailored to the task at hand (quantile estimation) and still provides greater numerical efficiency. It is of course possible for the RNE to be larger than 1.

6 Frequentist coverage and sensitivity

We now turn to the determination of the frequentist coverage of upper limits calculated with the correlated prior method of section 3.4. For this we adopt the framework of reference [9] and start by specifying an auxiliary experiment to measure the acceptance A. Since the acceptance is between 0 and 1, a good model is the binomial one. We imagine testing the acceptance cuts on k events, m of which pass. The likelihood of this auxiliary experiment is:

$$\mathcal{L}(A \,|\, m) \;=\; \binom{k}{m} A^{m}(1-A)^{k-m}. \tag{6.1}$$

Next, we select a uniform prior for A, $\pi(A) = 1$, and calculate the posterior:

$$p(A \,|\, m) \;=\; \frac{(k+1)!}{m!\,(k-m)!}\; A^{m}(1-A)^{k-m}, \tag{6.2}$$

which is a beta distribution. Its mean and standard deviation are:

$$\bar{A} \;=\; \frac{m+1}{k+2}, \tag{6.3a}$$

$$\sigma_A \;=\; \frac{1}{k+2}\sqrt{\frac{(m+1)\,(k+1-m)}{k+3}} \;=\; \sqrt{\frac{\bar{A}\,(1-\bar{A})}{k+3}}. \tag{6.3b}$$

Note that $\bar{A}$ is different from the maximum likelihood estimate m/k due to the contribution of prior information. If no data is taken, we have $m = k = 0$ and $\bar{A} = 1/2$, $\sigma_A = 1/\sqrt{12}$, in agreement with expectations from a uniform prior.

The above posterior will now be used as the acceptance prior in the main experiment, whose likelihood is given by equation (3.1). We also use the correlated prior of equation (3.10), with a prior of the form $\mu_S^{-\gamma}$ for the number of signal events μ_S. Putting all these components together, the posterior for the main experiment is:

$$p(\sigma, A \,|\, n) \;\propto\; \frac{(\sigma A L + bL)^{n}}{n!}\; e^{-\sigma A L - bL}\; (\sigma A L)^{-\gamma}\; AL\; \frac{(k+1)!}{m!\,(k-m)!}\; A^{m}\,(1-A)^{k-m}. \tag{6.4}$$

Since we are interested in upper limits on σ, we need to normalize the right-hand side of this expression to 1, integrate it over A, and produce the cumulative marginal distribution for σ. After some simplifications, this yields:

$$\begin{aligned} I\!P(\sigma \,|\, n) \;&\equiv\; \int_0^{\sigma} d\sigma' \int_0^1 dA\; p(\sigma', A \,|\, n) \\ &=\; \frac{\displaystyle\int_0^{\sigma L} dt\; \frac{(t+bL)^{n}}{n!}\; e^{-t}\; t^{-\gamma}\; I_{1-t/\sigma L}(k-m+1, m+1)}{\displaystyle\sum_{i=0}^{n} \frac{(bL)^{i}}{i!}\; \frac{\Gamma(n-i-\gamma+1)}{\Gamma(n-i+1)}}, \end{aligned} \tag{6.5}$$

where $I_x(a,b)$ is the incomplete beta function:

$$I_x(a,b) \;\equiv\; \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} \int_0^x dt\; t^{a-1}\,(1-t)^{b-1}. \tag{6.6}$$

Upper limits with credibility level α are obtained by solving $I\!P(\sigma_{ul}\,|\,n) = \alpha$ for σ_{ul}. According to equation (6.5), these upper limits depend on n, m, k, L, b, γ, and α, but for simplicity we will suppress the dependencies on L, b, γ and α in the equations that follow. The coverage of the procedure used to calculate upper limits depends on the true values σ_0 and A_0 of the signal cross section and acceptance, and is given by the sum:

$$C(\sigma_0, A_0) \;=\; \sum_{\substack{n,m \text{ such that}\\ \sigma_{ul}(n,m,k)\geq\sigma_0}} \frac{(\sigma_0 A_0 L + bL)^n}{n!}\, e^{-\sigma_0 A_0 L - bL} \binom{k}{m} A_0^m\,(1-A_0)^{k-m}. \tag{6.7}$$

This calculation can be simplified by noting the following properties of the upper limit:

1. $\sigma_{ul}(n,m,k)$ increases with increasing n (the number of events observed in the primary experiment);

2. $\sigma_{ul}(n,m,k)$ decreases with increasing m (the number of events observed in the auxiliary experiment).

These properties are fairly intuitive but they can also be proved rigorously. Because of the first property, for any given σ_0, m, and k there exists a smallest integer $\tilde{n}(\sigma_0,m,k)$ for which $\sigma_{ul}(\tilde{n}(\sigma_0,m,k),m,k) \geq \sigma_0$. In other words, $\tilde{n}(\sigma_0,m,k)$ is the smallest outcome of the experiment that yields *actual* coverage.[4] An equivalent definition of $\tilde{n}(\sigma_0,m,k)$ is that it is the smallest integer for which $I\!P(\sigma_0\,|\,\tilde{n}(\sigma_0,m,k)) \leq \alpha$ (this second inequality is easier to verify numerically than the first one). The coverage can then be rewritten as:

$$\begin{aligned} C(\sigma_0, A_0) \;&=\; \sum_{m=0}^{k} \binom{k}{m} A_0^m\,(1-A_0)^{k-m} \sum_{n=\tilde{n}(\sigma_0,m,k)}^{\infty} \frac{(\sigma_0 A_0 L + bL)^n}{n!}\, e^{-\sigma_0 A_0 L - bL}, \\ &=\; \sum_{m=0}^{k} \binom{k}{m} A_0^m\,(1-A_0)^{k-m}\, P(\tilde{n}(\sigma_0,m,k), \sigma_0 A_0 L + bL), \end{aligned} \tag{6.8}$$

where we used the well-known relationship between the tail of the Poisson distribution and the incomplete gamma function $P(n,x)$, equation (3.24).

The coverage is a highly discontinuous function of the true cross section σ, but a continuous function of the acceptance A. The explanation for this is similar to that provided in [9, §4.3]. To deal graphically with discontinuities, we simply evaluate the coverage at 1000 different points and plot them as single dots without connecting

[4] In the usual frequentist definition, coverage is an *average* over an ensemble of experiments and is a number between 0 and 1. On the other hand, *actual* coverage refers here to whether or not the extracted upper limit σ_{ul} exceeds the true value σ_0 in a particular experiment, and is *either* 0 *or* 1.

them. Figures 12 and 13 show some coverage curves as a function of cross section and acceptance.

For the $\gamma = 0$ curves, there appears to be undercoverage whenever the relative uncertainty on the expected number of signal events in the primary experiment is smaller than the relative uncertainty on the acceptance in the auxiliary experiment. Consider for example the $L = 100$ pb^{-1} curve in the upper left plot of Figure 12. That curve crosses the 90% coverage line at a cross section of about 2.5 pb, which corresponds to $2.5 \times 100 \times 0.1 = 25$ signal events. The relative uncertainty on 25 signal events is 20%, which is also the expected relative uncertainty on an acceptance measurement with $A_0 = 0.1$ and $k = 225$: $\Delta A/A_0 = \sqrt{(1 - A_0)/(kA_0)} = 0.2$. A similar calculation can be made for the $L = 100$ pb^{-1} curve in the upper right plot of Figure 12.

Undercoverage appears to be more pronounced for the $\gamma = 1/2$ curves.

6.1 Large-sample behavior of coverage

We just showed that, in the framework of reference [9], upper limits derived from the correlated prior method do not cover for all values of the parameters. It is therefore interesting to study what happens in the limit of a large data sample, since in that situation the likelihood is approximately Gaussian, and it is known that for a flat prior, Bayesian upper limits on the location parameter of a Gaussian have exact coverage. We first investigate what is meant by a large data sample in a Poisson measurement, where one only has a single "data point", namely the number of events n. Suppose we repeat the measurement k times and obtain a sequence $\{n_1, \ldots, n_k\}$ of event counts. The combined likelihood for the k measurements is then:

$$\begin{aligned}
\mathcal{L}(\sigma, A \,|\, \{n_1, \ldots, n_k\}) &= \prod_{i=1}^{k} \left\{ \frac{(\sigma A L + bL)^{n_i}}{n_i!} \, e^{-\sigma A L - bL} \right\}, \\
&= \frac{[kL(\sigma A + b)]^N}{N!} \, e^{-kL(\sigma A + b)} \;\; \frac{N!}{n_1! \cdots n_k!} \left(\frac{1}{k}\right)^{n_1} \cdots \left(\frac{1}{k}\right)^{n_k}, \\
&\propto \frac{[kL(\sigma A + b)]^N}{N!} \, e^{-kL(\sigma A + b)}, \qquad (6.9)
\end{aligned}$$

where $N \equiv \sum_{i=1}^{k} n_i$ and we dropped a factor independent of the likelihood parameters at the third line. This little calculation demonstrates that repeating k times a measurement with integrated luminosity L is equivalent to performing the measurement once with an integrated luminosity of $k \times L$. The integrated luminosity is therefore the parameter that goes to infinity in the asymptotic limit.

Figure 14 shows how the coverage of the upper limits varies with integrated luminosity L. Interestingly, the coverage does not seem to approach the nominal value of 90% as L increases, for either choice of prior ($\gamma = 0$ or $\gamma = 1/2$). We note however, that the coverage improves if *both* the integrated luminosity L of the main experiment and the sample size k of the auxiliary experiment are increased. This can be understood as follows. In the way we modeled our problem, the likelihood is given by:

$$\mathcal{L}(\sigma, A \,|\, n) = \frac{(\sigma A L + bL)^n}{n!} \, e^{-\sigma A L - bL}, \qquad (6.10)$$

and depends on two unknown parameters, σ and A. The maximum likelihood estimator of σ is not unique since the data n only depends on the product of σ and A, and A is unknown. This lack of uniqueness of the maximum likelihood estimator invalidates a standard theorem of Bayesian statistics, according to which Bayesian upper limits have coverage with a precision of order $\mathcal{O}(n^{-1/2})$, where n is the sample size (or the integrated luminosity in our case).[18]

On the other hand, if we combine the primary and auxiliary experiment likelihoods:

$$\mathcal{L}(\sigma, A \mid n, m) \;=\; \frac{(\sigma A L + bL)^n}{n!}\, e^{-\sigma A L - bL} \binom{k}{m} A^m (1-A)^{k-m}, \tag{6.11}$$

then the combined data (n, m) determine a unique maximum likelihood estimator for σ. As a result, the Bayesian coverage theorem will apply in the limit where both k and L become very large.

One way to model an increase in k is to treat it as an observation from a normalization process that is Poisson with mean μL, where L is the same integrated luminosity as in the primary experiment, and μ, the cross section times acceptance of the normalization process, is assumed known exactly. The combined likelihood becomes then:

$$\mathcal{L}(\sigma, A \mid n, m) \;=\; \frac{(\sigma A L + bL)^n}{n!}\, e^{-\sigma A L - bL}\, \frac{(\mu L)^k}{k!}\, e^{-\mu L} \binom{k}{m} A^m (1-A)^{k-m}. \tag{6.12}$$

It is easy to verify that the posterior distribution derived from this likelihood is identical to equation (6.5), provided that the same priors are used for σ and A. Upper limits will therefore also be identical. The coverage calculation is different however, since we now have to sum over the observation k. Instead of equation (6.7), we have:

$$C(\sigma_0, A_0) \;=\; \sum_{\substack{n,m,k \text{ such that} \\ \sigma_{ul}(n,m,k) \geq \sigma_0}} \frac{(\sigma_0 A_0 L + bL)^n}{n!}\, e^{-\sigma_0 A_0 L - bL}\, \frac{(\mu L)^k}{k!}\, e^{-\mu L} \binom{k}{m} A_0^m\, (1-A_0)^{k-m}. \tag{6.13}$$

Figure 15 shows some examples of coverage curves versus integrated luminosity for this model. In all cases, the actual coverage approaches its nominal value as L becomes very large.

Another feature demonstrated by Figure 15 is that the $\gamma = 0$ prior seems to yield faster convergence of the coverage to its nominal value than the $\gamma = 1/2$ prior. In general, not all priors yield the same rate of convergence and it is possible to use such considerations to optimize prior selection. Reference [17] shows that for a problem with two parameters, one of interest (σ) and one nuisance (A), it is possible to choose a prior for which the coverage will match its nominal value with a precision of $\mathcal{O}(n^{-1})$ instead of $\mathcal{O}(n^{-1/2})$. Such a prior is called "coverage-matching." To construct it, one first seeks a reparametrization of the problem, $(\sigma, A) \to (\sigma, \lambda)$, where λ is orthogonal to σ in the sense of Fisher information: $E\left[\partial^2 \ln \mathcal{L} / \partial\sigma\partial\lambda\right] = 0$. The coverage-matching prior is then proportional to Jeffreys' prior for σ in the new parametrization. For the combined likelihood of equation (6.12), the parameter orthogonal to σ can be found by solving a differential equation and is given by:

$$\lambda \;=\; \sigma A - b\ln(\sigma A + b) - \mu\ln(1-A) \tag{6.14}$$

up to an additive constant. The coverage matching prior is then:

$$\begin{aligned}\pi(\sigma, A) &\propto\; g(\lambda(\sigma, A))\,\left\{E\left[-\left.\frac{\partial^2\ln\mathcal{L}}{\partial\sigma^2}\right|_{(\sigma,\lambda)}\right]\right\}^{\frac{1}{2}}\,\frac{\partial(\sigma,\lambda)}{\partial(\sigma, A)} \\ &\propto\; g(\lambda(\sigma, A))\,\frac{A}{1-A}\,\sqrt{\frac{\mu}{\sigma A + b}\left[\mu + A(1-A)\frac{\sigma^2}{\sigma A + b}\right]},\end{aligned} \tag{6.15}$$

where $g(\lambda)$ is an arbitrary function of λ. Unfortunately this prior is numerically not as tractable as the ones we studied previously.

6.2 Coverage with subjective nuisance priors

So far in our coverage calculations we assumed that information about the nuisance parameter A came from a single, well-defined auxiliary experiment. It was therefore possible to choose an objective prior for A in that auxiliary experiment, namely the uniform prior. Coverage was calculated with respect to an ensemble in which the data from both the primary and auxiliary experiments are fluctuated.

In a realistic application this will rarely be possible. A much more likely scenario is one where information about A comes from a combination of measurements, Monte Carlo calculations, beliefs grounded in theory, and so on. The resulting prior for A is then necessarily subjective (in the Bayesian sense), as explained in section 2.1. Accordingly, the setup we consider in this section consists of a single experiment with the likelihood of equation (3.1) and a correlated prior for σ and A:

$$\pi(\sigma, A) \;=\; (\sigma A L)^{-\gamma}\, A L\, \pi(A), \tag{6.16}$$

$$\pi(A) \;=\; \frac{(k+1)!}{m!\,(k-m)!}\, A^m\,(1-A)^{k-m}. \tag{6.17}$$

This time however, we do not treat k and m as experimental data, but rather as parameters indexing a family of distributions. These parameters are assumed to have been chosen to best represent our prior state of knowledge about the acceptance. If for example we have estimates of the mean $\bar{A}$ and standard deviation σ_A, then equations (6.3) can be solved to yield the corresponding values of k and m. Since there are no longer any "auxiliary data" to fluctuate in the ensemble used to calculate coverage, we now study the question of how to handle the nuisance parameter in that calculation. Two approaches are possible: the coverage can be calculated at a number of *fixed* values of the nuisance parameter, or it can be *averaged* over the nuisance prior. We examine both approaches in the next couple of subsections.

6.2.1 Coverage at fixed values of the nuisance parameter

In this approach the nuisance parameter is treated on the same footing as the parameter of interest, and the coverage is checked on a grid of fixed values of both types of parameter. This is a strictly frequentist approach that treats the Bayesian upper limit calculation as a "black box" with no intrinsic merit. For our cross section measurement with acceptance uncertainties, the coverage at fixed values of the true cross section σ_0 and true acceptance A_0 is given by:

$$C(\sigma_0, A_0, L) \;=\; P(\tilde{n}(\sigma_0, m, k, L), \sigma_0 A_0 L + bL), \tag{6.18}$$

where P is the incomplete gamma function and $\tilde{n}(\sigma_0, m, k, L)$ is, as before, the smallest outcome of the experiment that yields *actual* coverage, i.e. for which the extracted upper limit on the cross section will *actually* exceed its true value σ_0.

Figure 16 shows $C(\sigma_0, A_0, L)$ as a function of integrated luminosity L and for $\sigma_0 =$ 20 pb and $A_0 = 0.15$, 0.1, 0.0785, and 0.05. The acceptance prior has a mean of 0.10 and a standard deviation of 0.02. As long as the true acceptance is approximately equal to, or larger than, the mean prior acceptance, coverage will eventually reach 100% at large integrated luminosity. To understand better the coverage behavior, we return to equation (6.18). Generically, the incomplete gamma function $P(a, x)$ is monotonic with respect to x and rises from 0 to 1 in a range of x centered on $a - 1$ and of width about $\sqrt{a}$.[19] Therefore, (over)coverage will be achieved whenever:

$$1 \;+\; (\sigma_0 A_0 + b)\, L \;\gg\; \tilde{n}(\sigma_0, m, k, L), \tag{6.19}$$

i.e. whenever the expected total number of events exceeds the minimal number of events needed to achieve actual coverage in a particular experiment. Both members of the above inequality depend on L. As shown in Figure 17, the dependence of $\tilde{n}(\sigma_0, m, k, L)$ on L is approximately linear. Therefore, whether or not coverage will be achieved in the asymptotic limit depends on whether the slope of $\tilde{n}$ vs. L is smaller or larger than $\sigma_0 A_0 + b$.

6.2.2 Coverage averaged over the nuisance prior

If we remember that the whole point of coverage is to provide a probability interpretation to measurement results, then from a Bayesian point of view it makes little sense to check coverage with respect to parameters with a subjective prior, because such parameters already have a properly defined probability measure prior to the measurement. With respect to parameters with an objective prior however, the situation is different: objective priors are often improper, in which case they do not define a probability measure. Coverage then provides a way to assign a probability interpretation to the resulting posterior.

It follows from the above that for problems involving a mixture of subjective and objective priors, it would be most useful to be able to decouple the effects of both types of prior on the coverage. As shown in appendix B of [9], Bayesian credible intervals do in fact have *average* frequentist coverage, where the average is calculated with respect to the prior density, assumed to be proper. Therefore, a simple way to decouple subjective and objective priors is to *average* the coverage over the subjective priors, and to check this average coverage at several fixed values of the parameter(s) with objective prior(s). In this regard it is interesting to quote from a recent report by professional statisticians on the subject of coverage matching priors, which suggests the above method for handling mixtures of subjective and objective priors, although in a more general context:

> *With the rapid advances in computational techniques for Bayesian statistics that exploit the increased computing power now available, researchers are able to adopt more realistic, and usually more complex, models. However, it is then less likely that the statistician will be able to properly elicit prior beliefs about all aspects of the model. Moreover, many parameters may not have a direct interpretation. This suggests that there is a need to develop general robust methods for prior specification that incorporate both subjective and nonsubjective components. In this case, the matching property could be recast as being the approximate equality of the posterior probability of a suitable set and the corresponding frequentist probability averaged over the parameter space with respect to any continuous prior that preserves the subjective element of the specified prior.* [18, pg. 27]

For the cross section measurement with acceptance uncertainties, this partially averaged coverage is given by:

$$C(\sigma_0, L) \;=\; \int_0^1 dA\, \frac{(k+1)!}{m!\,(k-m)!}\, A^m\, (1-A)^{k-m}\, P(\tilde{n}(\sigma_0, m, k, L), \sigma_0 AL + bL). \quad (6.20)$$

The coverage of 90% credibility level upper limits is shown as a function of integrated luminosity in Figure 18. One notes that the coverage tends to its nominal value as the integrated luminosity increases. Furthermore, only for the flat prior ($\gamma = 0$) does the coverage always remain *above* nominal.

6.3 Expected upper limits

Figure 19 shows the expected 90% credibility level upper limit as a function of true cross section and integrated luminosity. The expected upper limit is higher for 20% acceptance uncertainty ($k = 225$) than for zero acceptance uncertainty ($k = \infty$).

7 Summary

This note describes a fully Bayesian method for analyzing measurements of Poisson processes in the presence of acceptance and background uncertainties. We have introduced and motivated a prior that is uniform, i.e. "vague" in terms of information, in the number of signal events, rather than in the signal cross section as is usually done. The resulting posterior densities are always proper, even in the single-channel case, and can therefore be used to extract upper limits or two-sided intervals.

In discussing the numerical aspects of upper limit calculations, we have shown that the Monte Carlo importance sampling technique can be used very effectively in problems with large numbers of nuisance parameters. Asymptotic formulae for estimating uncertainties on these calculations are provided.

The frequentist properties of the method were studied for a variety of ensembles, some of which were motivated by a strict frequentist perspective on the problem, and some others by a more reasoned Bayesian perspective. Since the method itself is Bayesian, we favor the latter perspective. Accordingly, coverage should be averaged over the subjective prior(s), and this average coverage should be checked at several fixed values of the parameter(s) with objective prior(s). We noted that in the asymptotic limit (i.e. for large integrated luminosity) this coverage converges to its nominal value. In addition, when the prior for the number of signal events is taken to be uniform, the coverage always remains above nominal, ensuring the "conservatism" of upper limits calculated from small samples.

It is hoped that the methods described here will find widespread application and will supplant some of the more questionable, "ad hoc" procedures currently in use.

Acknowledgements

Many thanks to Louis Lyons for his detailed comments on a first draft of this note, and to Joel Heinrich for sharing his understanding of some of the coverage curves.

Appendix

A General formulation of the correlated prior method

A salient feature of the likelihood function (3.1) is that the data only depends on the *product* of the cross section and the acceptance. In this appendix we generalize this feature by considering measurements in which the data cannot distinguish between the parameter of interest θ and the nuisance parameter ν.[11] In mathematical terms, we assume that there exists a function $\eta(\theta, \nu)$, independent of the data x, such that the likelihood depends on θ and ν only through η:

$$\mathcal{L}(\theta, \nu \,|\, x) \;=\; \widetilde{\mathcal{L}}(\eta(\theta, \nu) \,|\, x). \tag{A.1}$$

If there is more than one parameter of interest or more than one nuisance parameter, we assume that η is a vector with the same dimension as θ, and that the Jacobian of the transformation $\theta \to \eta$ is non-singular.

The idea of the correlated prior method is to set up the measurement problem in such a way that nuisance parameters do *not* get updated by the measurement. Thus, given a nuisance prior $\pi(\nu)$, we search for a combined prior $\pi(\theta, \nu)$ such that:

$$\int d\theta \; \pi(\theta, \nu) \;=\; \pi(\nu) \qquad \text{and} \qquad p(\nu \,|\, x) \;=\; \pi(\nu). \tag{A.2}$$

With the help of Bayes' theorem, the equation on the right can be rewritten as:

$$\frac{\int d\theta \; \mathcal{L}(\theta, \nu \,|\, x) \, \pi(\theta, \nu)}{\int d\nu \int d\theta \; \mathcal{L}(\theta, \nu \,|\, x) \, \pi(\theta, \nu)} \;=\; \pi(\nu). \tag{A.3}$$

Next, we note that under a change of variable $\theta \to \eta \equiv \eta(\theta, \nu)$, probabilities remain invariant; in particular: $\pi(\theta, \nu)\; d\theta \;=\; \pi(\eta, \nu)\; d\eta$. Equation (A.3) can therefore be rewritten as:

$$\frac{\int d\eta \; \widetilde{\mathcal{L}}(\eta \,|\, x) \; \pi(\eta, \nu)}{\int d\nu \int d\eta \; \widetilde{\mathcal{L}}(\eta \,|\, x) \; \pi(\eta, \nu)} \;=\; \pi(\nu), \tag{A.4}$$

and is satisfied by any $\pi(\eta, \nu)$ that factorizes into $\pi(\eta)\, \pi(\nu)$. Transforming back to (θ, ν), the solution is:

$$\pi(\theta, \nu) \;=\; \pi(\eta(\theta, \nu)) \; \pi(\nu) \; \frac{\partial \eta}{\partial \theta}, \tag{A.5}$$

where $\partial\eta/\partial\theta$ is the Jacobian of the transformation $\theta \to \eta$. We will refer to the above solution as the correlated prior.

With this prior the marginal data density becomes:

$$p(x) \equiv \int d\theta \int d\nu \; \mathcal{L}(\theta, \nu \,|\, x) \; \pi(\theta, \nu), \tag{A.6}$$

$$= \int d\theta \int d\nu \; \widetilde{\mathcal{L}}(\eta(\theta, \nu) \,|\, x) \; \pi(\eta(\theta, \nu)) \; \pi(\nu) \; \frac{\partial \eta}{\partial \theta}, \tag{A.7}$$

$$= \int d\eta \int d\nu \; \widetilde{\mathcal{L}}(\eta \,|\, x) \; \pi(\eta) \; \pi(\nu), \tag{A.8}$$

$$= \int d\eta \; \widetilde{\mathcal{L}}(\eta \,|\, x) \; \pi(\eta). \tag{A.9}$$

Hence, by using the correlated prior we have "trivially" eliminated one integration from the calculation of $p(x)$, making it less likely to encounter divergent results when improper priors are used. Next, we write down an expression for the conditional posterior for the parameter of interest θ:

$$p(\theta \,|\, \nu, x) = \frac{p(\theta, \nu \,|\, x)}{p(\nu \,|\, x)} = \frac{1}{p(x)} \; \mathcal{L}(\theta, \nu \,|\, x) \; \pi(\eta(\theta, \nu)) \; \frac{\partial \eta}{\partial \theta}, \tag{A.10}$$

and for the marginal posterior for θ:

$$p(\theta \,|\, x) = \int d\nu \; p(\theta, \nu \,|\, x) = \int d\nu \; p(\theta \,|\, \nu, x) \; p(\nu \,|\, x) = \int d\nu \; p(\theta \,|\, \nu, x) \; \pi(\nu). \tag{A.11}$$

The marginal posterior for the parameter of interest is therefore equal to its conditional posterior averaged over the nuisance *prior*.

We emphasize that the correlated prior method uses Bayes' theorem and the rules of probability without any additional, external principle, other than a method for choosing a prior. It can be applied whenever the condition embodied in eq. (A.1) is satisfied.

B Search for Poisson processes in two or more independent channels

In this appendix we demonstrate that the usual way of factorizing the prior, in terms of cross section and acceptance, leads to a proper posterior when the search is performed in two or more channels with independent acceptances. For two channels, the likelihood is:

$$\mathcal{L}(\sigma, A_1, A_2 \,|\, n_1, n_2) = \frac{(\sigma A_1 L + b_1 L)^{n_1}}{n_1!} \; \frac{(\sigma A_2 L + b_2 L)^{n_2}}{n_2!} \; e^{-\sigma A_1 L - b_1 L - \sigma A_2 L - b_2 L}, \tag{B.1}$$

and the joint prior for the three parameters (σ, A_1, A_2) is:

$$\pi(\sigma, A_1, A_2) = G_1(A_1) \; G_2(A_2), \quad \text{where } G_i(A) \equiv \frac{e^{-\frac{1}{2}\left(\frac{A - A_{i0}}{\Delta A_i}\right)^2}}{\sqrt{2\pi} \, K_i \, \Delta A_i}, \tag{B.2}$$

and K_1, K_2 are normalization constants. This prior is improper with respect to σ. To prove that the posterior is nevertheless proper, we must show that the marginal data density:

$$p(n_1, n_2) \;=\; \int_0^{+\infty} d\sigma \int_0^1 dA_1 \int_0^1 dA_2 \; \mathcal{L}(\sigma, A_1, A_2 \,|\, n_1, n_2)\; \pi(\sigma, A_1, A_2) \tag{B.3}$$

is finite, i.e. neither zero nor infinite. Inserting the expressions for $\mathcal{L}(\sigma, A_1, A_2 \,|\, n_1, n_2)$ and $\pi(\sigma, A_1, A_2)$ into $p(n_1, n_2)$, and expanding the power factors in the likelihood yields:

$$\begin{aligned} p(n_1, n_2) \;=\; & \sum_{i=0}^{n_1}\sum_{j=0}^{n_2} \frac{1}{i!\,j!}\,\frac{(b_1 L)^{n_1-i}}{(n_1-i)!}\,\frac{(b_2 L)^{n_2-j}}{(n_2-j)!}\; e^{-b_1 L - b_2 L} \\ & \int_0^{+\infty} d\sigma \int_0^1 dA_1 \int_0^1 dA_2 \;\, (\sigma A_1 L)^i \;\, (\sigma A_2 L)^j \;\, e^{-\sigma A_1 L - \sigma A_2 L}\; G_1(A_1)\; G_2(A_2). \end{aligned} \tag{B.4}$$

The integral over σ can now be performed:

$$\begin{aligned} p(n_1, n_2) \;=\; & \sum_{i=0}^{n_1}\sum_{j=0}^{n_2} \frac{(i+j)!}{i!\,j!}\,\frac{(b_1 L)^{n_1-i}}{(n_1-i)!}\,\frac{(b_2 L)^{n_2-j}}{(n_2-j)!}\; e^{-b_1 L - b_2 L} \\ & \frac{1}{L}\int_0^1 dA_1 \int_0^1 dA_2 \; \frac{A_1^i\; A_2^j}{(A_1 + A_2)^{i+j+1}}\; G_1(A_1)\; G_2(A_2). \end{aligned} \tag{B.5}$$

Finally, we do a transformation on the integration variables: $(A_1, A_2) \to (r, \phi)$, with

$$\begin{cases} A_1 \equiv r^2\; (\sin\phi)^2, \\ A_2 \equiv r^2\; (\cos\phi)^2. \end{cases} \tag{B.6}$$

The Jacobian of this transformation is:

$$\frac{\partial(A_1, A_2)}{\partial(r, \phi)} \;=\; 2\; r^3\; \sin(2\phi). \tag{B.7}$$

Since the integration region is a square in (A_1, A_2) space, it is not trivial to write it in terms of r and ϕ. This is not necessary however, because all we need is a finite *upper bound* on $p(n_1, n_2)$, and this can be obtained even if we integrate over the *whole* quadrant $A_1 \geq 0$, $A_2 \geq 0$. If we do that, equation (B.5) implies the following inequality:

$$\begin{aligned} p(n_1, n_2) \;\leq\; & 4 \sum_{i=0}^{n_1}\sum_{j=0}^{n_2} \frac{(i+j)!}{i!\,j!}\,\frac{(b_1 L)^{n_1-i}}{(n_1-i)!}\,\frac{(b_2 L)^{n_2-j}}{(n_2-j)!}\; e^{-b_1 L - b_2 L} \\ & \frac{1}{L}\int_0^{\pi/2} d\phi \int_0^{+\infty} dr \;\; r \;\; (\sin\phi)^{2i+1} \;\; (\cos\phi)^{2j+1} \;\; G_1(r^2 (\sin\phi)^2)\; G_2(r^2 (\cos\phi)^2). \end{aligned} \tag{B.8}$$

To verify that the integral converges, we divide the ϕ integration region in two sections, from 0 to $\pi/4$ and from $\pi/4$ to $\pi/2$. For the first section, consider the following bound on the integrand:

$$\begin{aligned} r\,(\sin\phi)^{2i+1}\,(\cos\phi)^{2j+1}\,G_1(r^2(\sin\phi)^2)\,G_2(r^2(\cos\phi)^2) \\ &\leq r\,\cos\phi\,G_1(r^2(\sin\phi)^2)\,G_2(r^2(\cos\phi)^2) \\ &\leq r\,\cos\phi\,\frac{1}{\sqrt{2\pi}\,K_1\,\Delta A_1}\,G_2(r^2(\cos\phi)^2). \end{aligned}$$

The integral of the last expression on the right-hand side is clearly finite:

$$\int_0^{\pi/4} d\phi \int_0^{+\infty} dr\,\frac{r\,\cos\phi}{\sqrt{2\pi}\,K_1\,\Delta A_1}\,G_2(r^2(\cos\phi)^2) \;=\; \frac{1}{2\sqrt{2\pi}\,K_1\,\Delta A_1}\int_0^{\pi/4} d\phi\,\frac{1}{\cos\phi} \;<\; \infty,$$

and therefore the first section of the integral in (B.8) also converges. For the second section of that integral the argument is similar, but uses a different bound on the integrand:

$$\begin{aligned} r\,(\sin\phi)^{2i+1}\,(\cos\phi)^{2j+1}\,G_1(r^2(\sin\phi)^2)\,G_2(r^2(\cos\phi)^2) \\ &\leq r\,\sin\phi\,G_1(r^2(\sin\phi)^2)\,G_2(r^2(\cos\phi)^2) \\ &\leq r\,\sin\phi\,G_1(r^2(\sin\phi)^2)\,\frac{1}{\sqrt{2\pi}\,K_2\,\Delta A_2}. \end{aligned}$$

We have then:

$$\int_{\pi/4}^{\pi/2} d\phi \int_0^{+\infty} dr\,\frac{r\,\sin\phi}{\sqrt{2\pi}\,K_2\,\Delta A_2}\,G_1(r^2(\sin\phi)^2) \;=\; \frac{1}{2\sqrt{2\pi}\,K_2\,\Delta A_2}\int_{\pi/4}^{\pi/2} d\phi\,\frac{1}{\sin\phi} \;<\; \infty.$$

This ends the proof that $p(n_1, n_2)$ is finite in two dimensions. In higher dimensions one needs to introduce additional angles in the transformation (B.6), and the same result obtains.

C Asymptotic variance of upper limits estimated using importance sampling

Here we generalize a calculation described in section 2 of [20], and derive the asymptotic variance of quantiles estimated by importance sampling. Let f be a properly normalized density, of which we wish to estimate the α^{th} quantile θ_α, defined by the equation:

$$\int_{-\infty}^{\theta_\alpha} d\theta\,f(\theta) \;=\; \alpha, \quad 0 < \alpha < 1. \tag{C.1}$$

In a Bayesian context, with f a posterior density for θ, θ_α is the α credibility level upper limit on θ. Let $\hat{\theta}_\alpha$ be an estimate of θ_α obtained by a Monte Carlo calculation with importance sampling density g. Reference [20] gives the asymptotic variance of $\hat{\theta}_\alpha$ under the assumptions that the normalization of f is known a priori, and that there are no nuisance parameters. Neither of these two assumptions is satisfied by the upper limit calculations described in this note, and a generalization is therefore needed.

We start by defining a statistical functional T_α:

$$T_\alpha(H) \;\equiv\; W_H^{-1}(\alpha), \tag{C.2}$$

where $H(\theta, \vec{\nu})$ is an arbitrary probability distribution function depending on a parameter of interest θ and a vector of nuisance parameters $\vec{\nu}$, and:

$$W_H(t) \;\equiv\; \frac{\displaystyle\int r(\theta,\vec{\nu})\; \vartheta(t-\theta)\; dH(\theta,\vec{\nu})}{\displaystyle\int r(\theta,\vec{\nu})\; dH(\theta,\vec{\nu})} \tag{C.3}$$

for some nonnegative function $r(\theta, \vec{\nu})$. Let $F(\theta, \vec{\nu})$ and $G(\theta, \vec{\nu})$ be the cumulative probability distribution functions (cdf) corresponding, respectively, to the density of interest $f(\theta, \vec{\nu})$ and the importance sampling density $g(\theta, \vec{\nu})$. If we set $H = G$ and $r(\theta, \vec{\nu}) = f(\theta, \vec{\nu})/g(\theta, \vec{\nu})$ in the definition of $T_\alpha(H)$, we find:

$$T_\alpha(G) \;=\; W_G^{-1}(\alpha) \;=\; \theta_\alpha. \tag{C.4}$$

On the other hand, if $H = \widehat{G}$, the empirical distribution[5] of the Monte Carlo importance samples $\{(\theta_i, \vec{\nu}_i), i = 1, \ldots, n\}$, we have (still for $r = f/g$):

$$T_\alpha(\widehat{G}) \;=\; W_{\widehat{G}}^{-1}(\alpha) \;=\; \hat{\theta}_\alpha. \tag{C.5}$$

The statistical properties of $\hat{\theta}_\alpha$ can therefore be derived from those of the functional T_α. The latter can be studied with the help of its influence function, which we now calculate. If $H_\epsilon \equiv (1-\epsilon)H + \epsilon H_1$ for some cdf H_1, then:

$$W_{H_\epsilon}\left[W_{H_\epsilon}^{-1}(\alpha)\right] \;=\; \alpha. \tag{C.6}$$

Differentiating this equation with respect to ϵ in $\epsilon = 0$ yields:

$$\left.\frac{\partial W_{H_\epsilon}(y)}{\partial \epsilon}\right|_{\epsilon=0,\; y=W_{H_0}^{-1}(\alpha)} + \left.\frac{\partial W_{H_\epsilon}(y)}{\partial y}\right|_{\epsilon=0,\; y=W_{H_0}^{-1}(\alpha)} \left.\frac{\partial W_{H_\epsilon}^{-1}(\alpha)}{\partial \epsilon}\right|_{\epsilon=0} \;=\; 0. \tag{C.7}$$

This can be rearranged as follows:

$$\left.\frac{dT_\alpha(H_\epsilon)}{d\epsilon}\right|_{\epsilon=0} \;=\; -\,\frac{\partial W_{H_\epsilon}(y)/\partial\epsilon|_{\epsilon=0\; y=T_\alpha(H)}}{\partial W_{H_\epsilon}(y)/\partial y|_{\epsilon=0\; y=T_\alpha(H)}}. \tag{C.8}$$

[5]The *empirical* distribution of a sample of n data points assigns a probability of $1/n$ to each data point. It is discrete, and should be distinguished from the *parent* probability distribution of the sample, which it approximates.

The numerator and denominator on the right-hand side are straightforward to calculate; the result is:

$$\left.\frac{dT_\alpha(H_\epsilon)}{d\epsilon}\right|_{\epsilon=0} = \frac{\alpha \displaystyle\int_{-\infty}^{+\infty} d\theta \int d\vec{\nu}\; r(\theta,\vec{\nu})\; h_1(\theta,\vec{\nu}) \;-\; \int_{-\infty}^{T_\alpha(H)} d\theta \int d\vec{\nu}\; r(\theta,\vec{\nu})\; h_1(\theta,\vec{\nu})}{\displaystyle\int d\vec{\nu}\; r(T_\alpha(H),\vec{\nu})\; h(T_\alpha(H),\vec{\nu})}. \tag{C.9}$$

The influence function is now obtained by setting $h_1(\theta,\vec{\nu}) = \delta(\theta-\eta)\,\delta(\vec{\nu}-\vec{\mu})$ in the right-hand side. This function measures the influence on the functional T_α of a small proportion of observations at $(\eta,\vec{\mu})$ that do not come from H:

$$IF_{T_\alpha,H}(\eta,\vec{\mu}) \;=\; \frac{\alpha \;-\; \vartheta(T_\alpha(H)-\eta)}{\displaystyle\int d\vec{\nu}\; r(T_\alpha(H),\vec{\nu})\; h(T_\alpha(H),\vec{\nu})}\; r(\eta,\vec{\mu}). \tag{C.10}$$

Finally, we set $r(\eta,\vec{\mu}) = f(\eta,\vec{\mu})/g(\eta,\vec{\mu})$ and $h(\eta,\vec{\mu}) = g(\eta,\vec{\mu})$:

$$IF_{T_\alpha,G}(\eta,\vec{\mu}) \;=\; \frac{\alpha \;-\; \vartheta(\theta_\alpha-\eta)}{\displaystyle\int d\vec{\nu}\; f(\theta_\alpha,\vec{\nu})}\; \frac{f(\eta,\vec{\mu})}{g(\eta,\vec{\mu})}. \tag{C.11}$$

The usefulness of the influence function for our problem is that, as a consequence of the central limit theorem [21, section 6.3]:

$$\sqrt{n}\left[T_\alpha(\widehat{G}) - T_\alpha(G)\right] \;\to\; N\big(0,\gamma^2(G)\big) \tag{C.12}$$

as $n\to\infty$, where n is the number of samples used in the definition of $\widehat{G}$, and:

$$\gamma^2(G) \;=\; E_G\Big\{\big[IF_{T_\alpha,G}(\eta,\vec{\mu})\big]^2\Big\}. \tag{C.13}$$

Referring to equations (C.4) and (C.5), this result gives the asymptotic variance of the α credibility level upper limit $\hat{\theta}_\alpha$ as $\gamma^2(G)/n$, where $\gamma^2(G)$ is the expectation under G of the square of the influence function. Plugging expression (C.11) into (C.13) yields:

$$\begin{aligned} \gamma^2(G) \;&=\; \int_{-\infty}^{+\infty} d\eta \int d\vec{\mu}\; g(\eta,\vec{\mu}) \left[\frac{\alpha \;-\; \vartheta(\theta_\alpha-\eta)}{\displaystyle\int d\vec{\nu}\; f(\theta_\alpha,\vec{\nu})}\; \frac{f(\eta,\vec{\mu})}{g(\eta,\vec{\mu})}\right]^2 \\ &=\; \frac{1}{\left[\displaystyle\int d\vec{\nu}\; f(\theta_\alpha,\vec{\nu})\right]^2} \int_{-\infty}^{+\infty} d\eta \int d\vec{\mu}\; \frac{\Big\{\big[\alpha-\vartheta(\theta_\alpha-\eta)\big] f(\eta,\vec{\mu}\Big\}^2}{g(\eta,\vec{\mu})}, \end{aligned} \tag{C.14}$$

from which equation (5.9) in section 5.1 is derived.

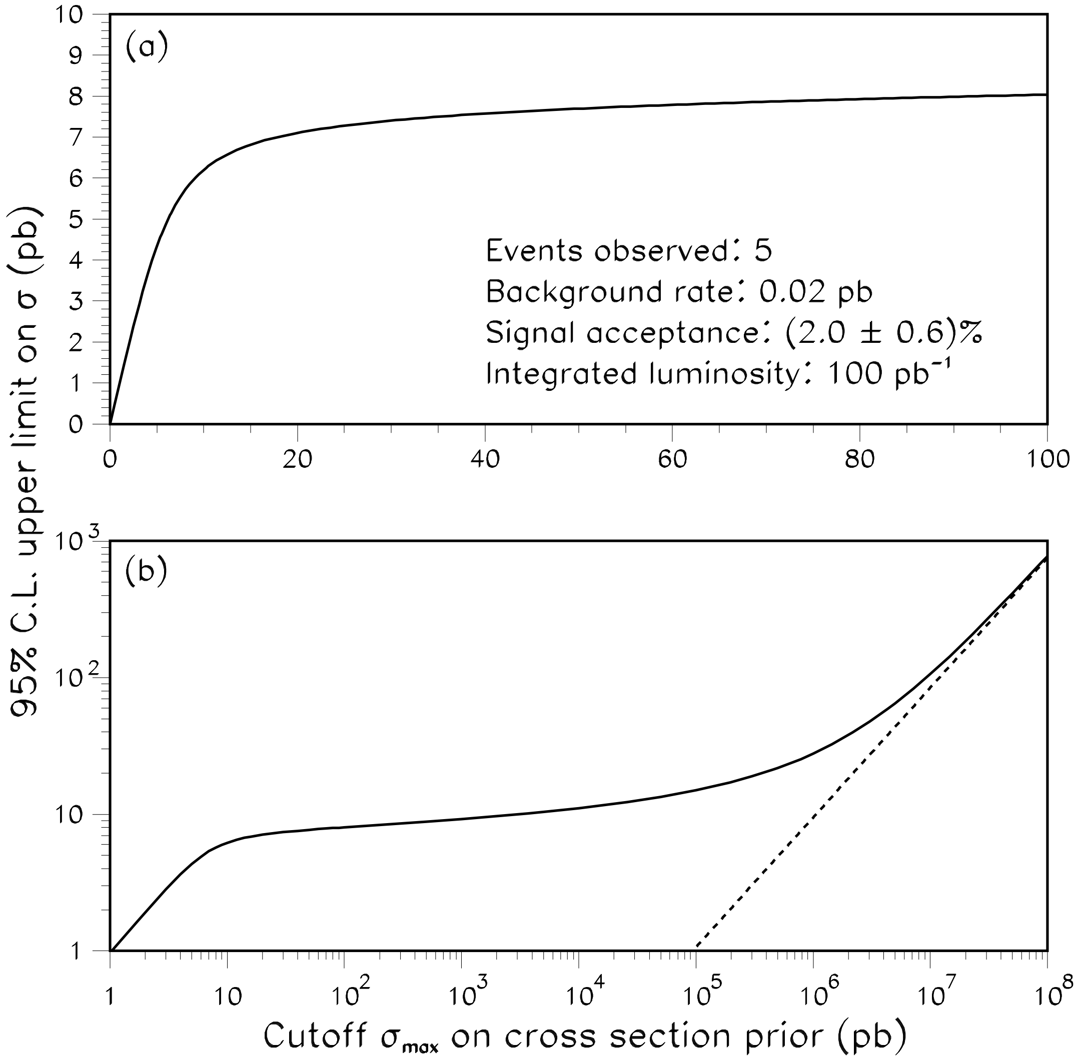

Figure 1: Bayesian upper limit at the 95% credibility level on a hypothetical cross section, as a function of the cutoff $\sigma_{\max}$ on the flat prior for that cross section. Plot (a) is linear and shows the upper limit variation for $\sigma_{\max}$ between 0 and 100 pb. Plot (b) is log-log and shows the effect of increasing $\sigma_{\max}$ up to 100 μb. The dashed line is an asymptote.

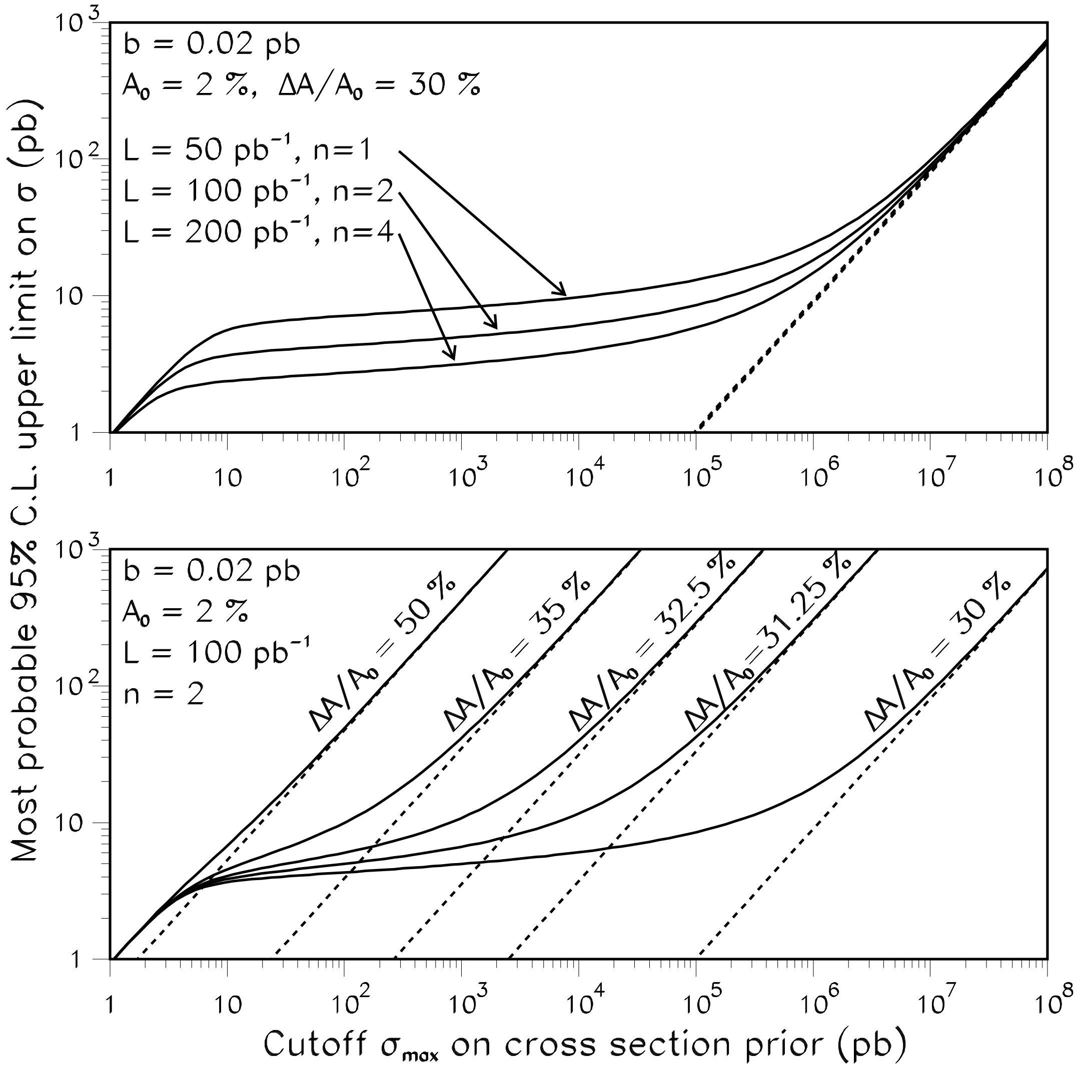

Figure 2: Upper limit at the 95% credibility level on a hypothetical cross section, as a function of the cutoff $\sigma_{\max}$ on the flat prior for that cross section. The observed number of events n is taken to be the most probable one for the given expected background rate b and integrated luminosity L, in the absence of signal. Top: the solid curves are for three different values of L and n. Bottom: the solid curves are for five different values of the relative acceptance uncertainty $\Delta A/A_0$. The dashed lines are asymptotes.

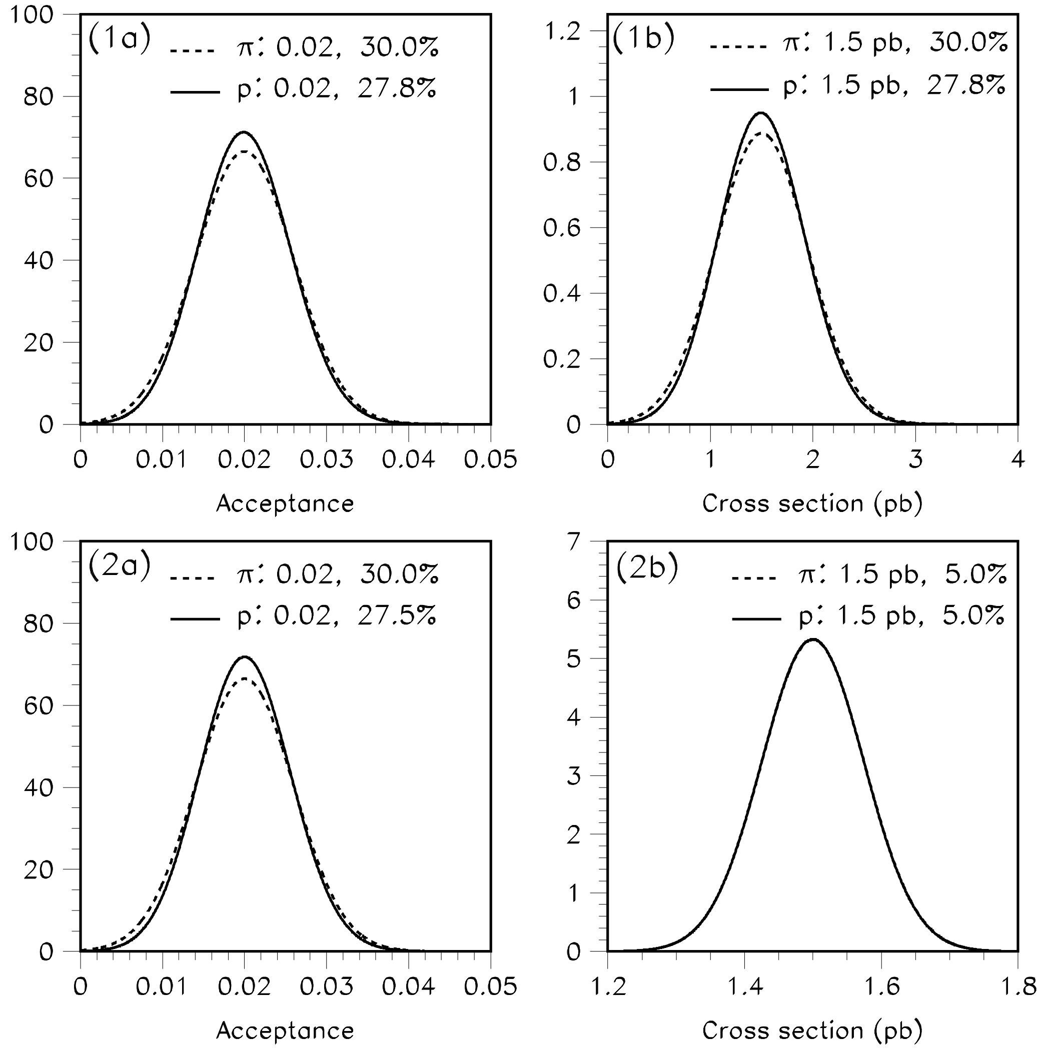

Figure 3: Two measurements in which the prior information is reinforced by the data. In both measurements, 5 events are observed over a background of 2.0. Measurement 1 (top) has a 30% prior uncertainty on both the acceptance and the cross section. Measurement 2 (bottom) has prior uncertainties of 30% on the acceptance and 5% on the cross section. The marginal prior and posterior densities are drawn with dashed and solid lines respectively, and their means and coefficients of variation ($\equiv$ RMS/mean) are indicated next to the symbols π and p. The marginal cross section posterior for the second measurement is indistinguishable from its prior.

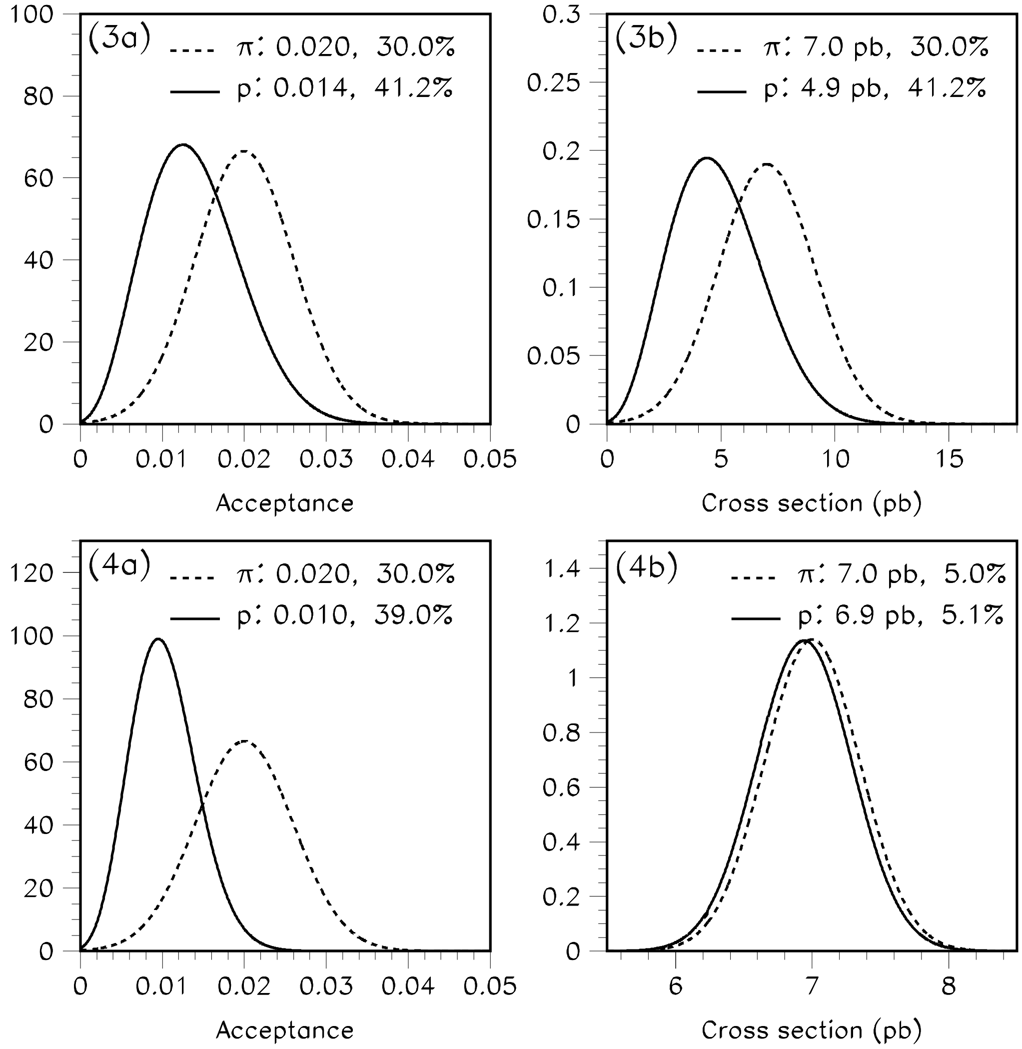

Figure 4: Two measurements in which the prior information conflicts with the data. In both measurements, 5 events are observed over a background of 2.0. Measurement 3 (top) has a 30% prior uncertainty on both the acceptance and the cross section. Measurement 4 (bottom) has prior uncertainties of 30% on the acceptance and 5% on the cross section. The marginal prior and posterior densities are drawn with dashed and solid lines respectively, and their means and coefficients of variation are indicated next to the symbols π and p.

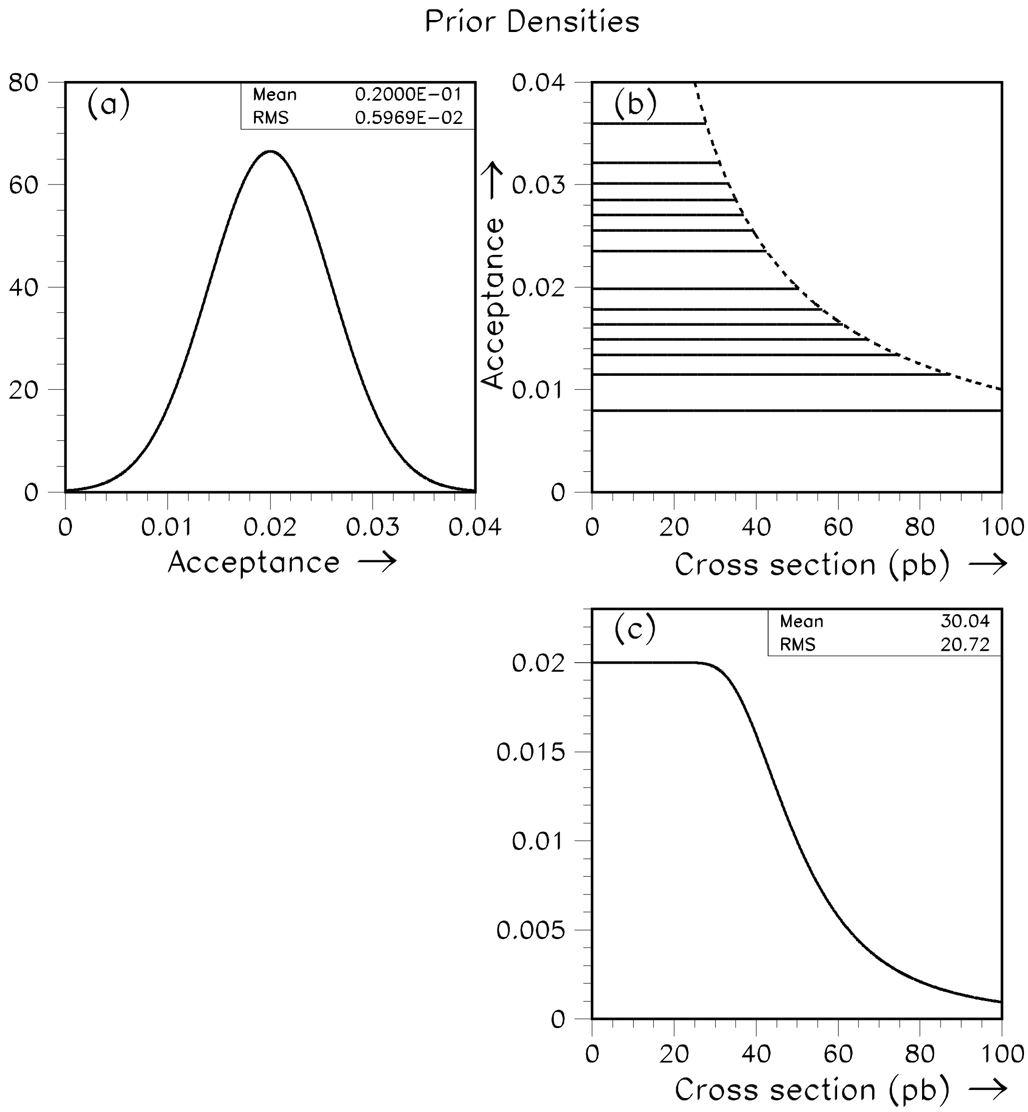

Figure 5: Prior probability density in the plane of acceptance versus cross section. The parameters of the prior distribution are: $A_0 = 0.020$, $\Delta A = 0.006$, $\mu_{\rm max} = 100$, and $L = 100\ {\rm pb}^{-1}$. Plot (b) shows contours of equal probability density (horizontal solid lines). The two outer lines correspond to the level of 5% of the maximum. Moving in, the contour levels increase in steps of 15% up to the 95% level for the two inner lines. The dashed line is not a contour but represents the $\mu_{\rm max}$ boundary: the probability density is zero above that line. Plot (a) shows the projection in acceptance and plot (c) that in cross section.

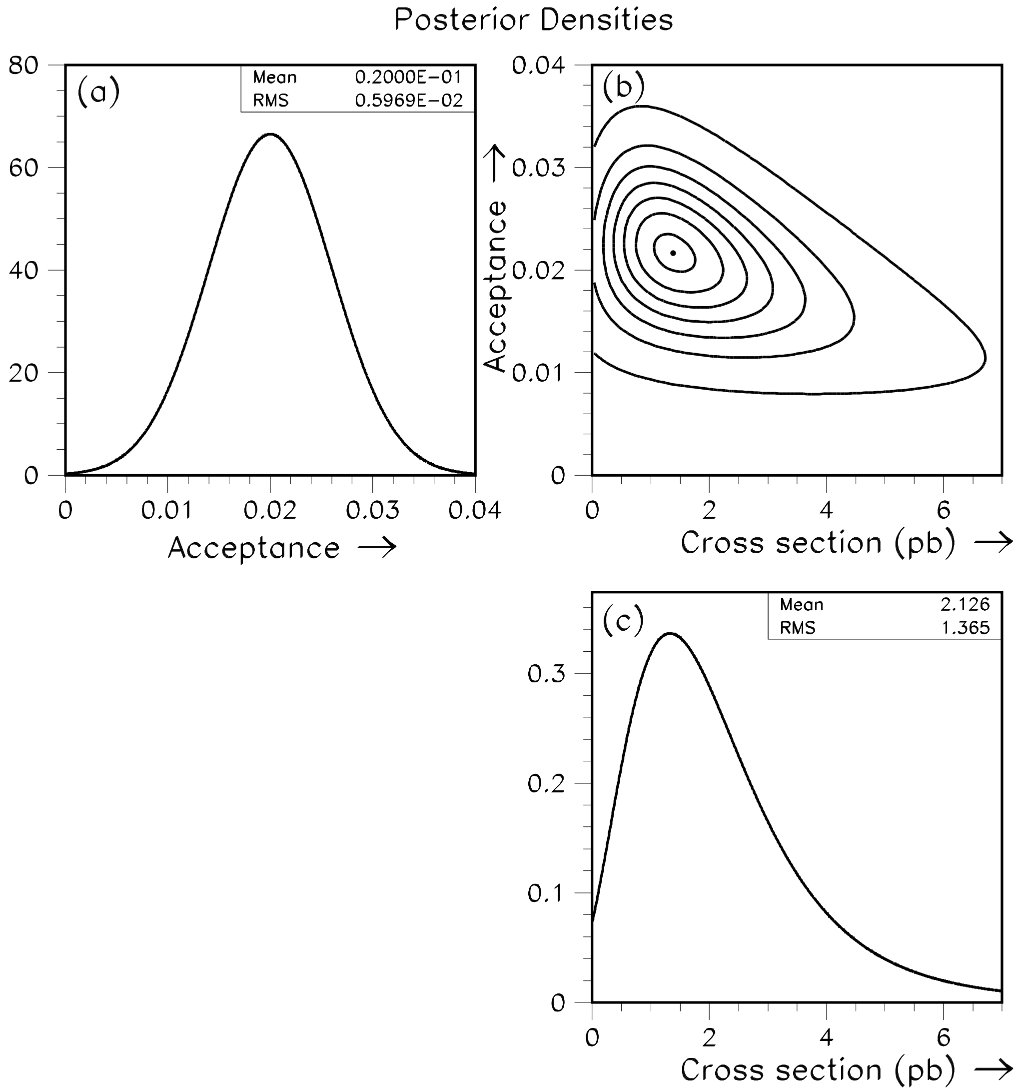

Figure 6: Posterior probability density in the plane of acceptance versus cross section, for the prior shown in Figure 5 and for a Poisson likelihood with $n = 5$ observed events and an expected background event rate of $b = 0.02$ pb. Plot (b) shows contours of equal probability density. The outer contour is at 5% of the maximum, and contour levels increase in steps of 15% as the maximum is approached. Plot (a) shows the projection in acceptance and plot (c) that in cross section. The marginal acceptance posterior is equal to the marginal acceptance prior by construction.

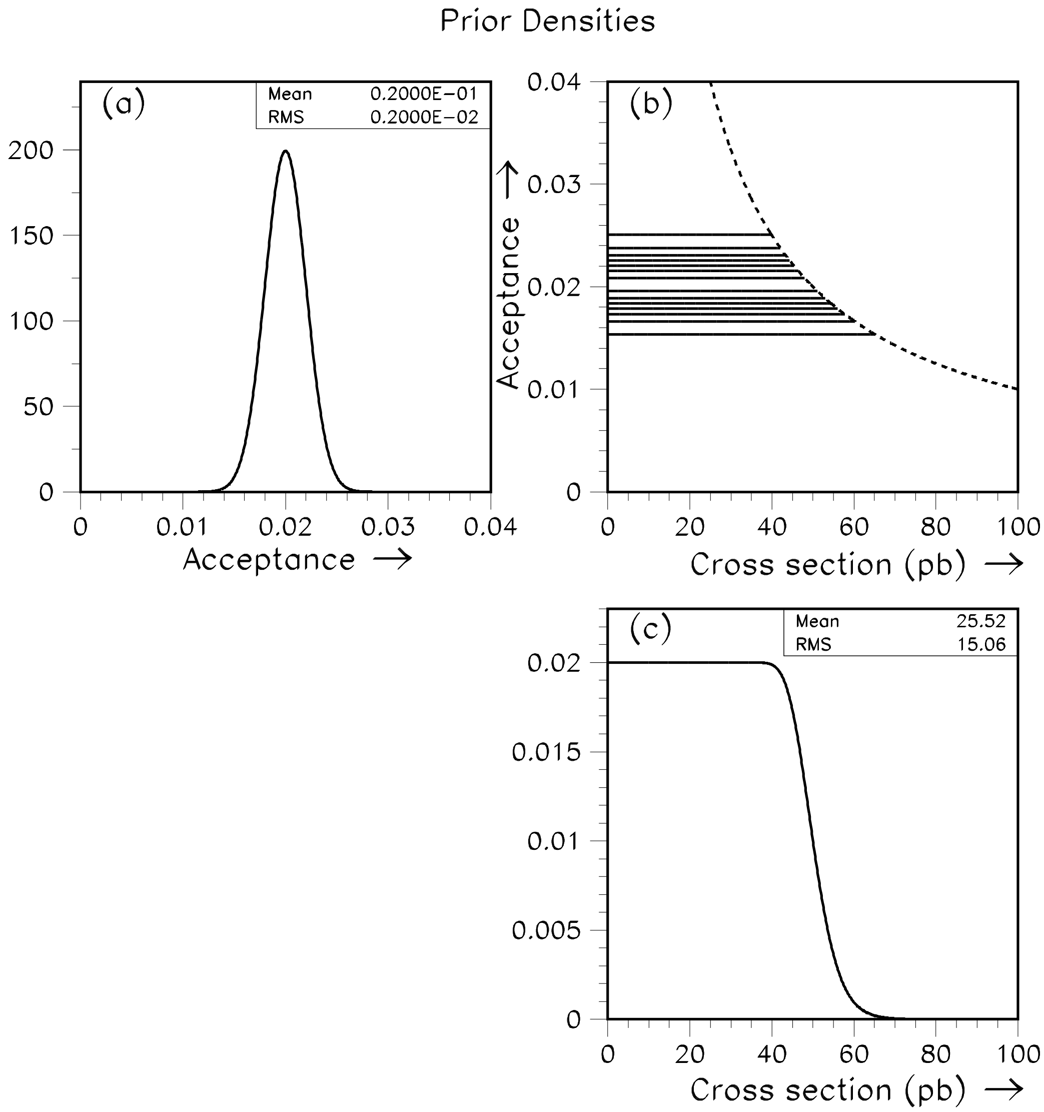

Figure 7: Same as Figure 5, but for $\Delta A = 0.002$, i.e. a narrower acceptance prior. The step in the cross section prior is correspondingly steeper (plot (c)).

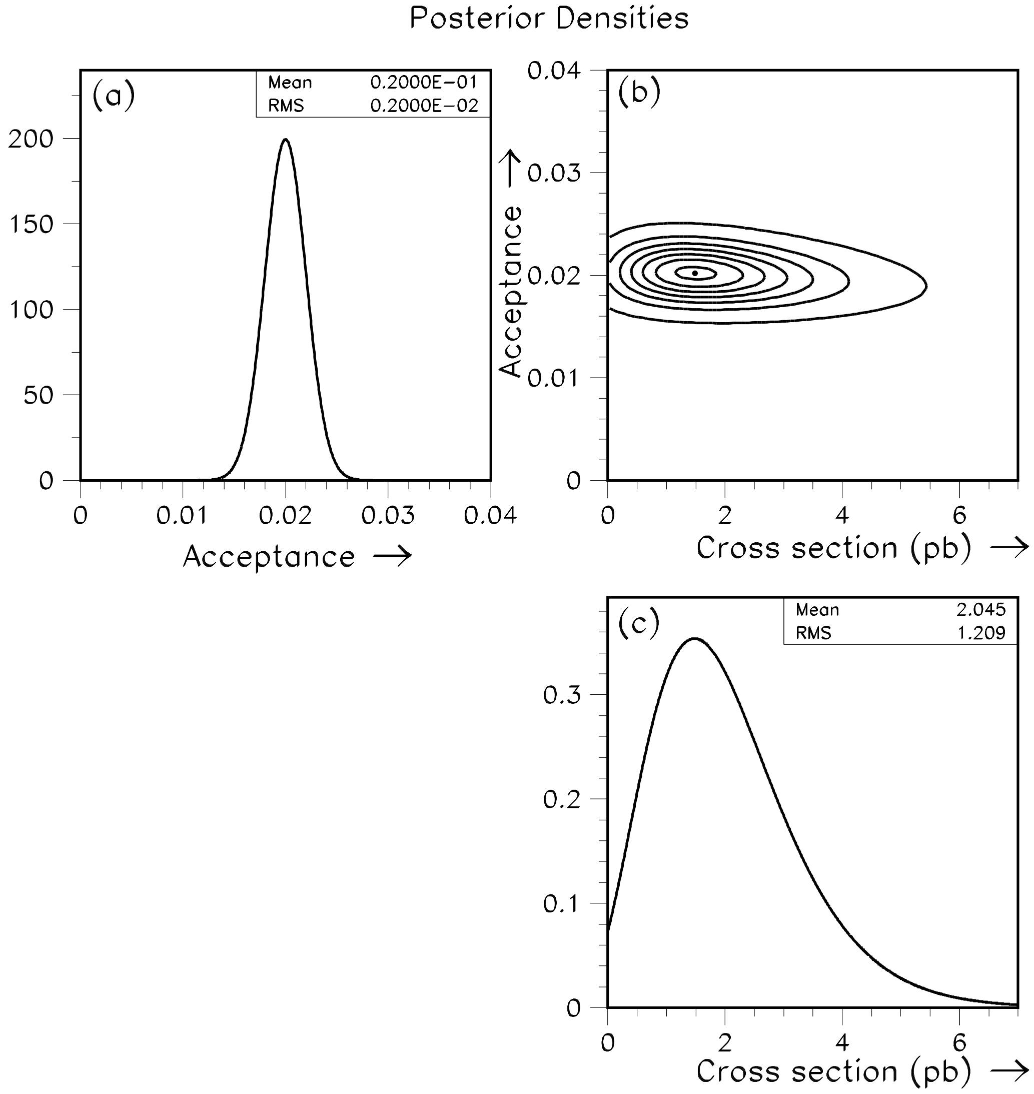

Figure 8: Same as Figure 6, but for $\Delta A = 0.002$, i.e. a narrower acceptance prior. Note that the posterior cross section density is correspondingly narrower (plot (c)).

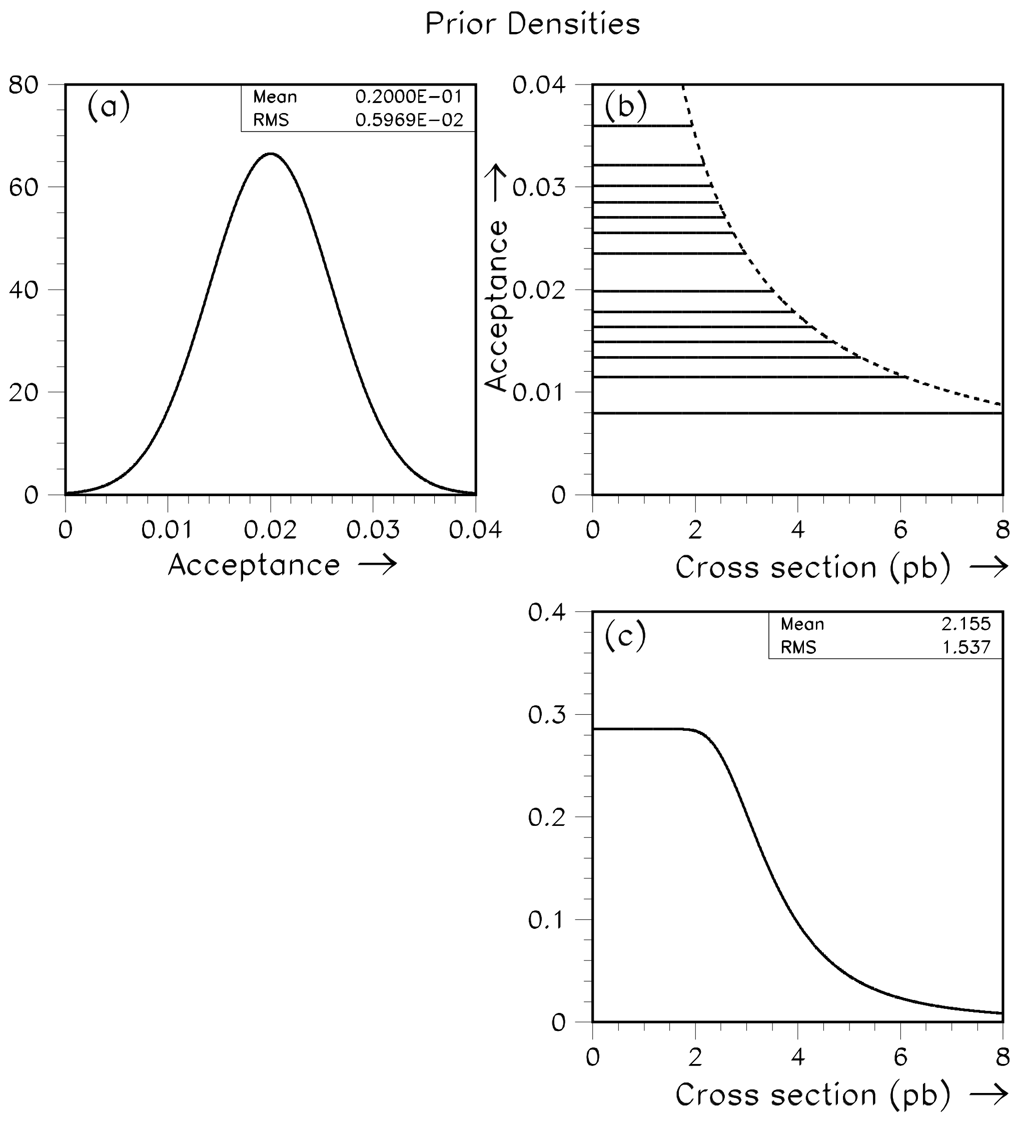

Figure 9: Same as Figure 5, but for $\mu_{\rm max} = 7$, i.e. a narrower cross section prior.

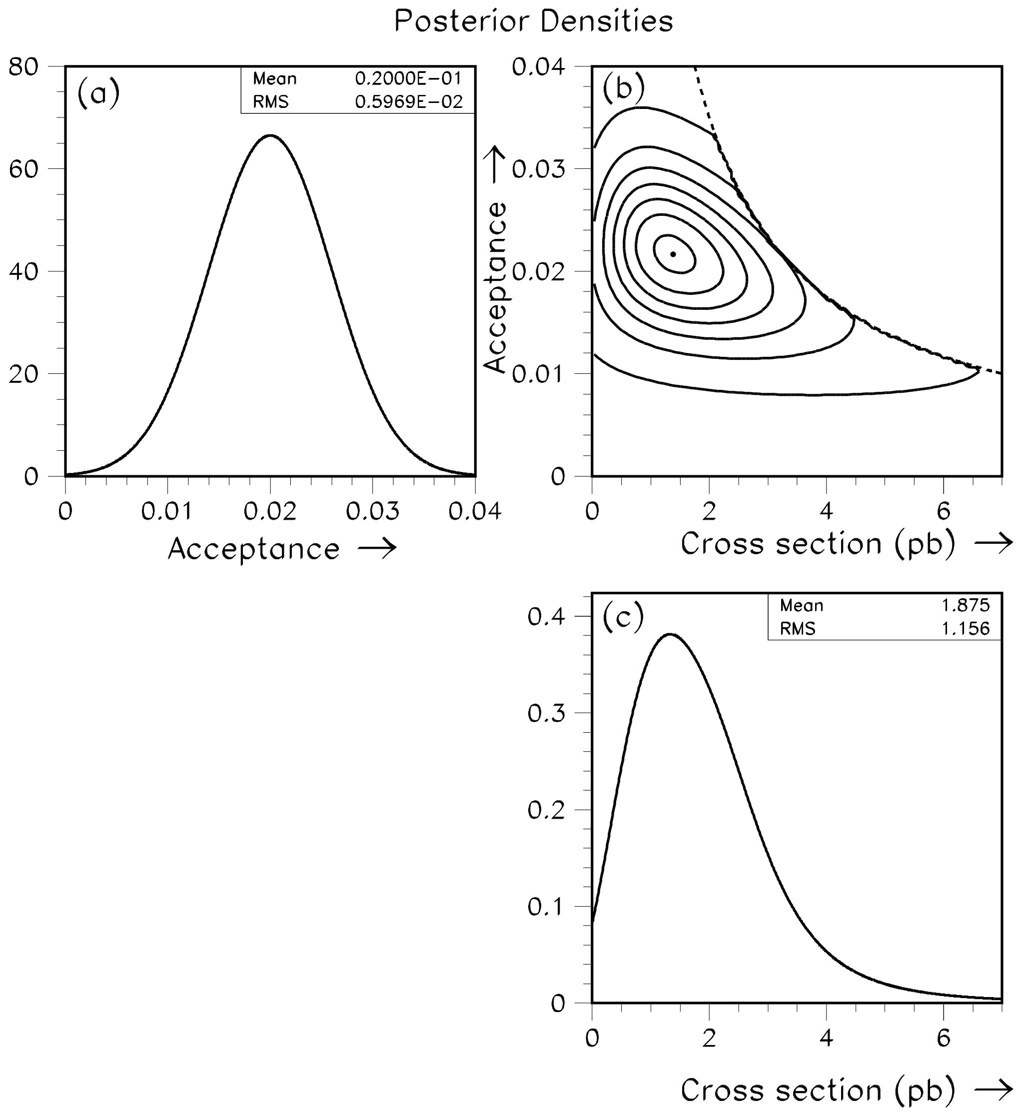

Figure 10: Same as Figure 6, but for $\mu_{\max} = 7$. The dashed line in plot (b) is not a contour but represents the $\mu_{\max}$ boundary, above which the posterior density is zero. Reducing $\mu_{\max}$ has not affected the acceptance posterior (by construction, see text), but has made the cross section posterior narrower.

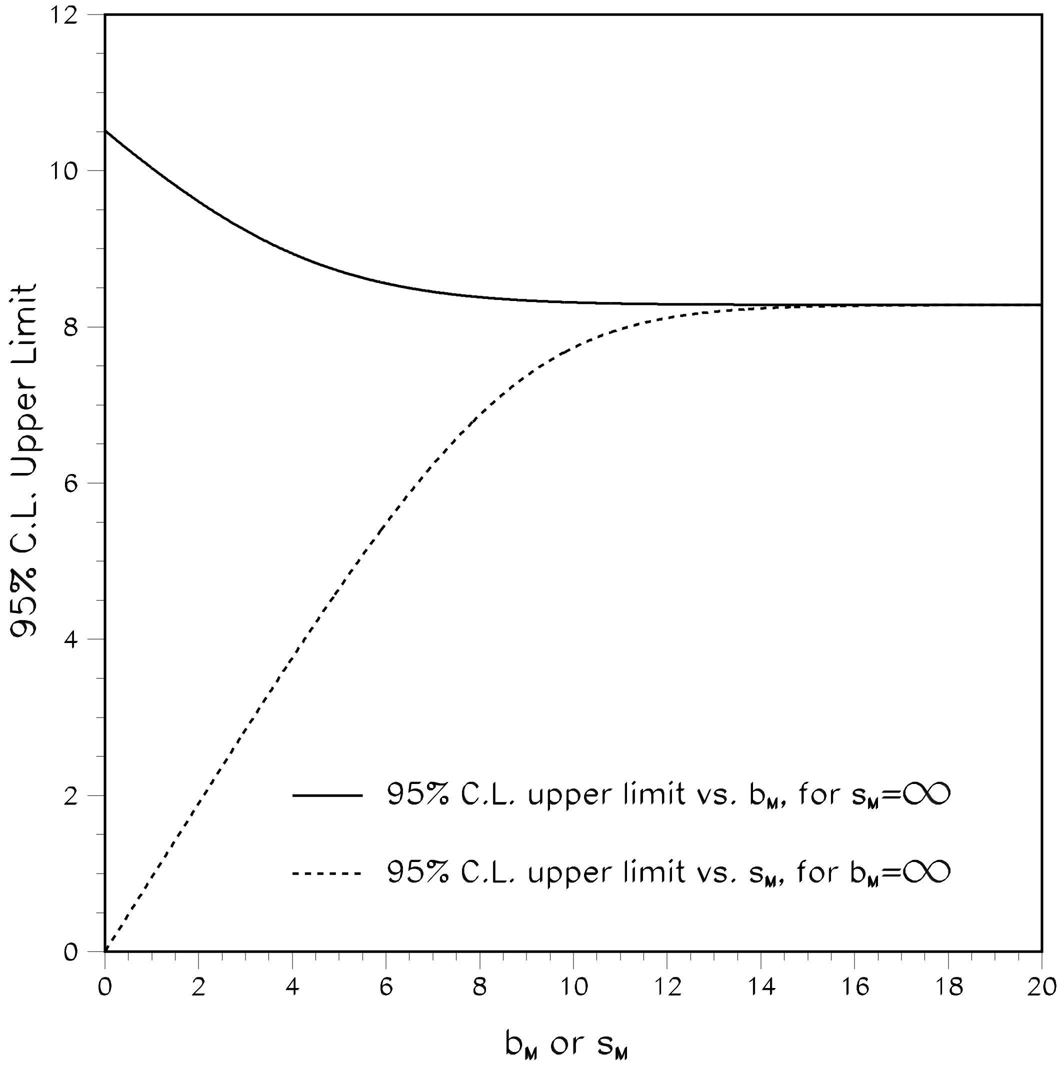

Figure 11: Upper limit at the 95% credibility level on the number of events from a Poisson signal process in the presence of background. The priors for the numbers of signal and background events are flat from 0 up to s_M and b_M, respectively. The observed number of events is 5. The solid line shows the upper limit as a function of b_M for $s_M = \infty$, whereas the dashed line shows the upper limit as a function of s_M for $b_M = \infty$. This plot demonstrates that extending a flat signal prior towards infinity has a conservative effect on upper limits, whereas extending a flat background prior is anti-conservative. Note how for small values of s_M the upper limit is practically equal to s_M, since in that case the prior information about the signal overwhelms the data.

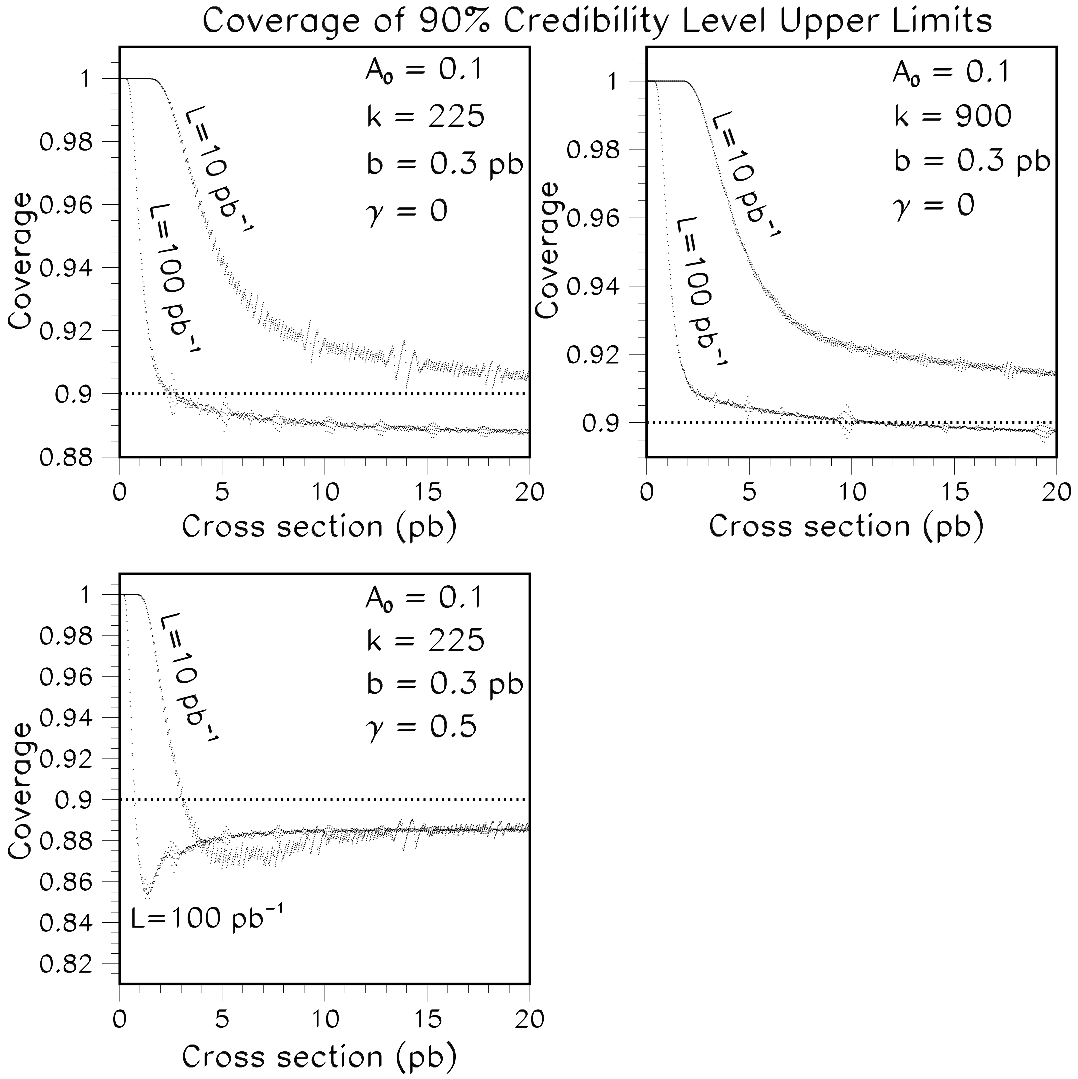

Figure 12: Coverage of 90% credibility level upper limits as a function of true cross section σ_0. For each plot, the true acceptance A_0 is indicated, as well as the total number of events k in the auxiliary experiment, the background b and the power γ of the signal prior. High values of k allow a more precise determination of the acceptance by the auxiliary experiment. The dotted lines indicate the credibility level of 90%.

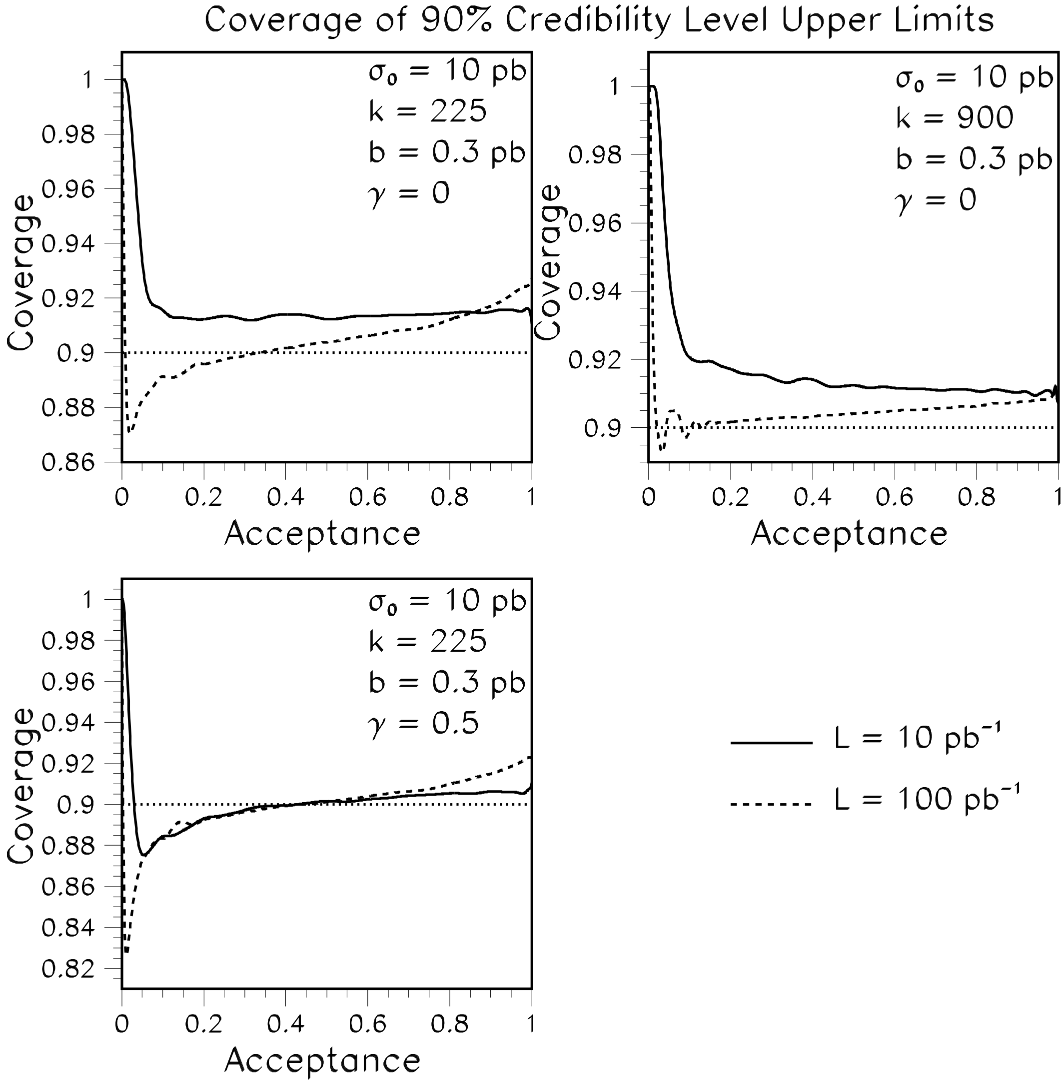

Figure 13: Coverage of 90% credibility level upper limits as a function of true acceptance A_0. For each plot, the true cross section σ_0 is indicated, as well as the total number of events k in the auxiliary experiment, the background b and the power γ of the signal prior. High values of k allow a more precise determination of the acceptance by the auxiliary experiment. The dotted lines indicate the credibility level of 90%.

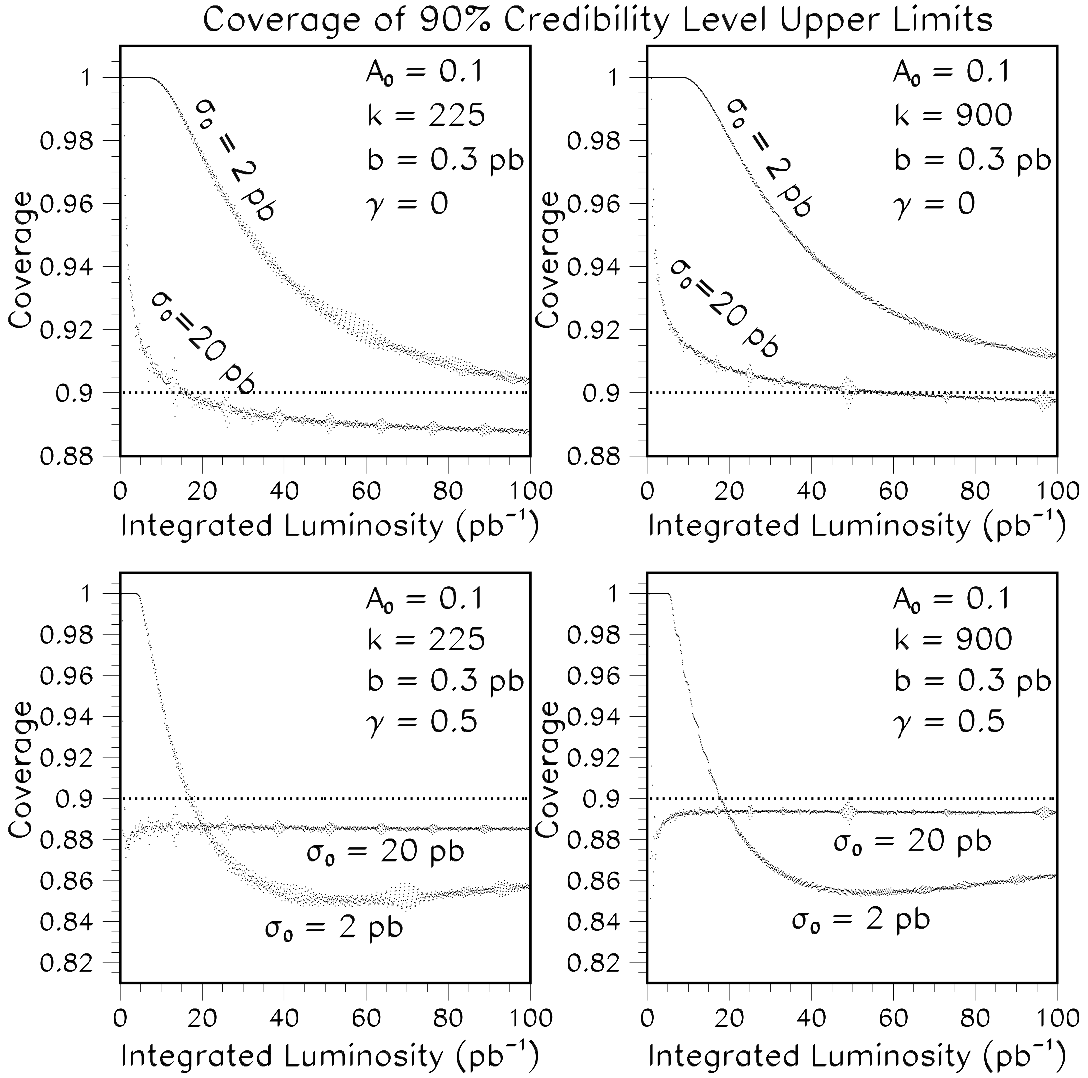

Figure 14: Coverage of 90% credibility level upper limits as a function of integrated luminosity L. For each plot, the true acceptance A_0 is indicated, as well as the total number of events k in the auxiliary experiment, the background rate b, and the power γ of the signal prior. In this version of the coverage calculation, the integrated luminosity only determines the event rate in the main experiment; the number of events in the auxiliary experiment is fixed at k. The dotted lines indicate the credibility level of 90%.

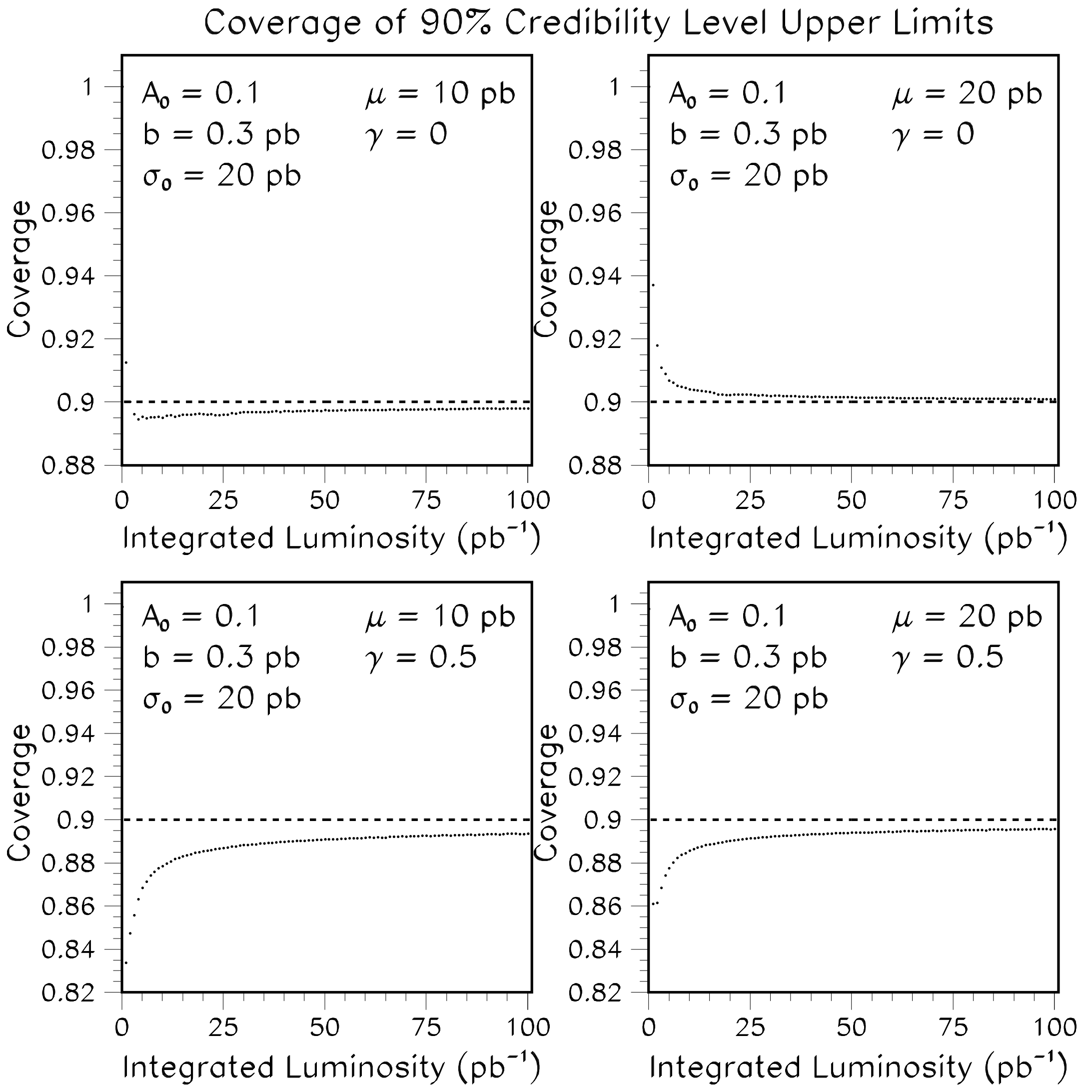

Figure 15: Coverage of 90% credibility level upper limits as a function of integrated luminosity L. For each plot, the true acceptance A_0 is indicated, as well as the event rate μ in the auxiliary experiment, the true signal cross section σ_0 and background rate b in the main experiment, and the power γ of the signal prior. In this version of the coverage calculation, the integrated luminosity determines both the event rate in the main experiment and in the auxiliary one. The dashed lines indicate the credibility level of 90%.

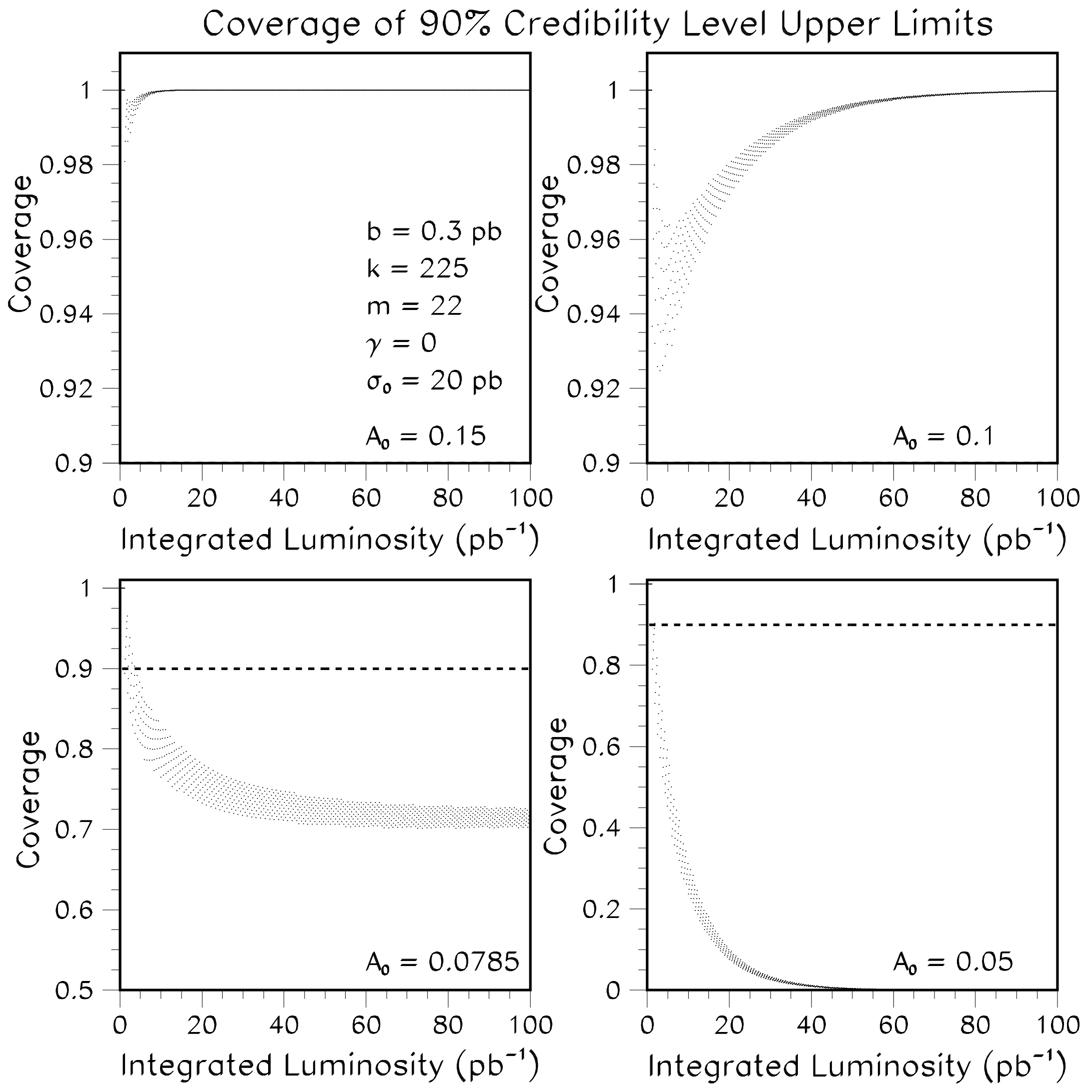

Figure 16: Coverage of 90% credibility level upper limits as a function of integrated luminosity L, for four different values of the true signal acceptance A_0. In the ensemble used to calculate coverage, only the number of events observed in the primary experiment is fluctuated. The values of the parameters b, k, m, γ, and σ_0 (see text) are the same in all four plots and are listed in the top left one. The dashed lines indicate the credibility level of 90%. Note the different Y-axis scales.

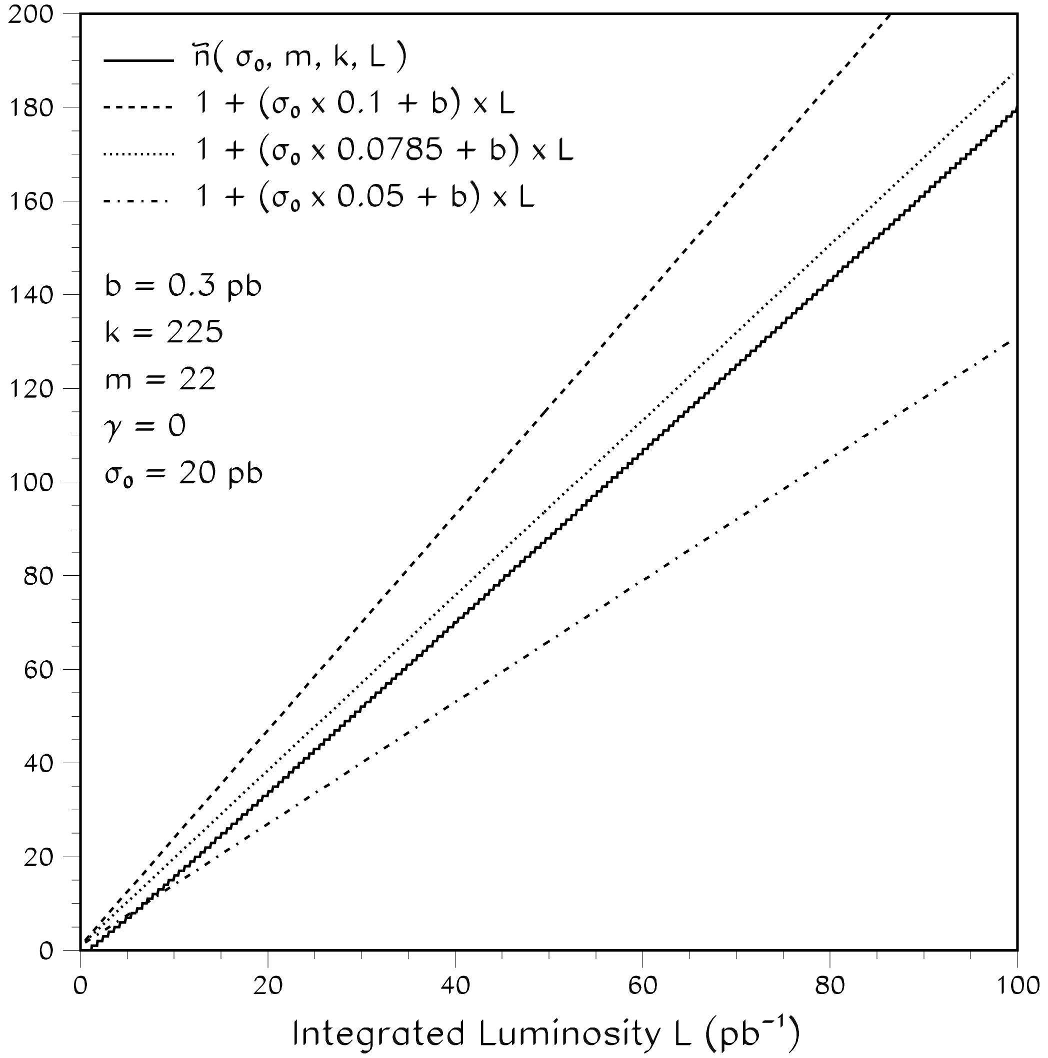

Figure 17: Minimum number of events required to achieve actual coverage in a particular experiment, $\tilde{n}(\sigma_0, m, k, L)$, as a function of integrated luminosity L (solid line). This is compared with the expected total number of events, plus one event, for three values of the true acceptance: $A_0 = 0.1$ (dashed line), $A_0 = 0.0785$ (dotted line), and $A_0 = 0.05$ (dot-dashed line). The credibility level α is set at 90%.

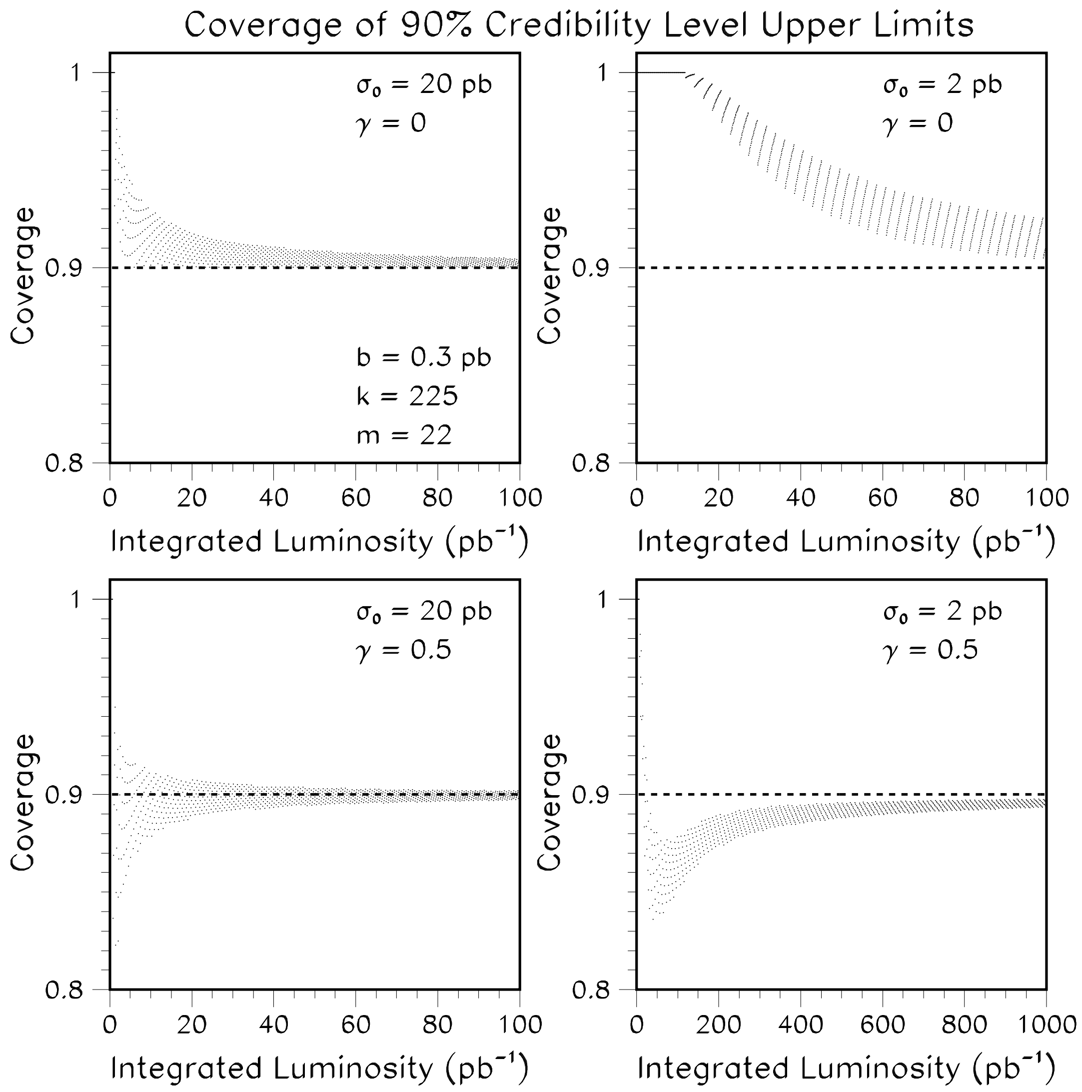

Figure 18: Coverage of 90% credibility level upper limits as a function of integrated luminosity L, for different values of the true signal cross section σ_0 and the power γ of the signal prior. In the ensemble used to calculate coverage, only the number of events observed in the primary experiment is fluctuated, and the coverage is then averaged over the acceptance prior. The values of the parameters b, k, and m (see text) are the same in all four plots and are listed in the top left one. The dashed lines indicate the credibility level of 90%. Note the extended X-axis scale in the bottom right plot.

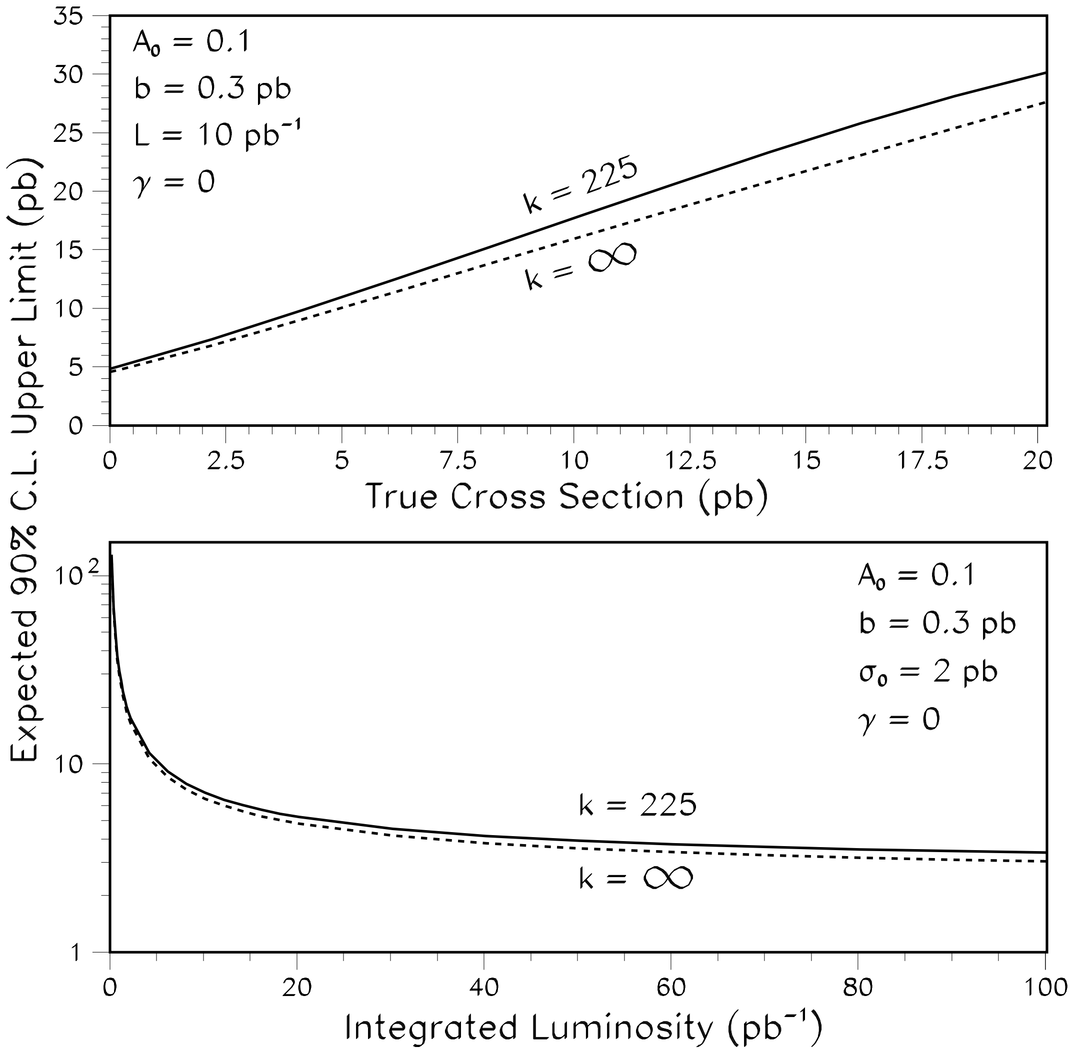

Figure 19: Expected 90% credibility level upper limits as a function of true cross section (top) and integrated luminosity (bottom). The relative acceptance uncertainty is 20% for the solid lines ($k = 225$) and 0 for the dashed lines ($k = \infty$).

References

[1] I. J. Good, "The Bayes/Non-Bayes Compromise: A Brief Review," J. Amer. Statist. Assoc. **87**, 597 (1992).

[2] F. Abe *et al.* (CDF Collaboration), "Top-quark search in the electron+jets channel in proton-antiproton collisions at $\sqrt{s} = 1.8$ TeV," Phys. Rev. D **43**, 664 (1991).

[3] Dennis V. Lindley, "The 1988 Wald Memorial Lectures: The Present Position in Bayesian Statistics," Statist. Sci. **5**, 44 (1990).

[4] Luc Demortier, "Constructing Ensembles of Pseudo-Experiments," in Proceedings of *PHYSTAT2003: Statistical Problems in Particle Physics, Astrophysics and Cosmology*, Stanford Linear Accelerator Center, Stanford, California, September 8-11, 2003, eConf C030908, SLAC-R-703; see also arXiv:physics/0312100 v2 (16 Dec 2003).

[5] Robert E. Kass and Larry Wasserman, "The Selection of Prior Distributions by Formal Rules," J. Amer. Statist. Assoc. **91**, 1343 (1996).

[6] E. T. Jaynes, "Probability Theory: The Logic of Science," edited by G. Larry Bretthorst, Cambridge University Press, 2003 (727pp).

[7] A. P. Dawid, M. Stone, and J. V. Zidek, "Critique of E. T. Jaynes's 'Paradoxes of Probability Theory'," University College London Department of Statistical Science Research Report 172, September 13, 1996; `http://www.ucl.ac.uk/Stats/research/Resrprts/psfiles/172.zip`.

[8] Ruoyong Yang and James O. Berger, "A Catalog of Noninformative Priors," ISDS discussion paper 97-42 (1997); `http://ftp.isds.duke.edu/WorkingPapers/97-42.ps`.

[9] Joel Heinrich, Craig Blocker, John Conway, Luc Demortier, Louis Lyons, Giovanni Punzi, and Pekka K. Sinervo, "Interval estimation in the presence of nuisance parameters; 1. Bayesian approach," CDF note 7117 (September 27, 2004); arXiv:physics/0409129 v1.

[10] Anthony O'Hagan, "Bayesian Inference," vol. 2B of "Kendall's Advanced Theory of Statistics," 1$^{\text{st}}$ edition, Halsted Press, New York, 1994 (330pp).

[11] Luc Demortier, "Bayesian treatments of systematic uncertainties," Proceedings of the Conference on *Advanced Statistical Techniques in Particle Physics*, Institute for Particle Physics Phenomenology, University of Durham, UK, 18-22 March 2002; IPPP/02/39 (2002), pg. 145.

[12] J. Linnemann, "Upper Limits and Priors," talk at the *Fermilab Workshop on Confidence Limits*, March 2000, `http://conferences.fnal.gov/cl2k/`.

[13] Jose M. Bernardo, "Reference Posterior Distributions for Bayesian Inference," J. R. Statist. Soc. B **41**, 113 (1979).

[14] D. R. Cox and N. Reid, "Parameter Orthogonality and Approximate Conditional Inference," J. R. Statist. Soc. B **47**, 1 (1987).

[15] M. Corradi, "Inclusion of systematic uncertainties in upper limits and hypothesis tests," in *Proceedings of the CERN Workshop on Confidence Limits, 17-18 January 2000*, CERN Yellow Report 2000-005.
[This paper provides an example of the use of the correlated prior method in High-Energy Physics. Unfortunately, its derivation of the method is confusing and incorrect: equation (1) gives the marginal posterior for the interest parameter as $\int d\vec{x}\; g(\vec{x})\; f_t'(t\,|\,\vec{x})$, where $f_t'(t\,|\,\vec{x})$ is the conditional posterior for the interest parameter t given the nuisance parameters $\vec{x}$, and $g(\vec{x})$ is the prior for the nuisance parameters; for that equation to be correct from the point of view of probability theory, $g(\vec{x})$ should be the *posterior* for the nuisance parameters, not their *prior*, as pointed out in [11].]

[16] John Geweke, "Bayesian inference in econometric models using Monte Carlo integration," Econometrica **57**, 1317 (1989).

[17] Robert Tibshirani, "Noninformative priors for one parameter of many," Biometrika **76**, 604 (1989).

[18] Gauri Sankar Datta and Trevor J. Sweeting, "Probability matching priors," Research Report No. 252, Department of Statistical Science, University College London (March 2005); `http://www.ucl.ac.uk/Stats/research/Resrprts/psfiles/rr252.pdf`.

[19] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling, "Numerical Recipes; The art of scientific computing," Cambridge University Press, 1986 (818pp).

[20] M. Vernon Johns, "Importance Sampling for Bootstrap Confidence Intervals," J. Amer. Statist. Assoc. **83**, 709 (1988).

[21] E. L. Lehmann, "Elements of Large-Sample Theory," Springer-Verlag, New York, Inc., 1999 (631pp).